

New crisis for British research

British civil research seems to be in for yet another crisis, this time because the government is bent on level budgeting. The time has come for radical defensive measures.

THE newest crisis in the support of basic research in Britain appears to have blown up more suddenly than seemed possible only a week ago. The Advisory Board for the Research Councils, which is to publish today last year's advice to the Secretary of State for Education and Science on the size of the science budget, will be found to have asked for more — and to have been refused. Now, with a new annual budget cycle about to begin, the board has no choice but to ask again. Like *Oliver Twist*, it is almost certain again to be refused. And then there will be unpalatable choices to be made by everybody.

Unpalatable choices? Surely those have already been made? That is what the principals in this endless and sometimes tedious saga will say. The trouble is that all those concerned have different perceptions of the problem. The central question is how much money the British Government should transfer each year to the research councils, the chief sources of support for academic research in Britain, and to less costly institutions such as the Royal Society and the British Museum (Natural History). The government considers it has done well by these organizations, having stood by the promise of the Prime Minister, Mrs Margaret Thatcher, four years ago that the science budget would be "protected" and seeing no return for its fortitude in a dramatic improvement of economic prospects. The research councils nevertheless consider that they have been unfairly squeezed, some of them because of the rapid increase of the cost of subscriptions to international laboratories such as the European Organization for Nuclear Research (CERN) in Geneva, others because they are being forced to choose between supporting university-based research and that carried out within their own institutes. But the ultimate beneficiaries of the science budget, university and laboratory scientists, sense at first hand the steady decline in the sparkle of the British research enterprise (see *Nature* 15 March, p.214), note the frequency with which even good research proposals are now turned down by the research councils, worry about the future of the institutions in which they happen to work — and become infected by the accidie now rife.

Conventions

These three orthogonal views of what has gone wrong will not be reconciled, let alone made the basis for a remedy, unless the pall of convention that hangs over everything can be dispelled. So it is worthwhile asking what would be done for British research if the existing administrative machinery were scrapped, and rebuilt from scratch. Even in shoestring Britain, the amounts of money spent are quite considerable. In the financial year about to end, the research councils and their smaller partners will have spent more than £600 million, one-sixth of that supplied by government departments in recompense for work undertaken on contract. The numbers of scientists and engineers teaching in British universities exceed 16,000 but not by much, and there are probably as many others working in research council institutes, as post-doctoral fellows and in polytechnics with pretensions to research. What this implies is that the income of the British research enterprise probably works out at an average of £20,000 a year for each active investigator — not an ungenerous allowance by ordinary standards, especially because some members of the research community make only modest demands of the system (or even none).

So why does so much money so constantly disappoint the reci-

pipients? Because much of it never gets to them. The British research enterprise is not just a machine for spending money on research projects but a huge establishment in its own right. The research councils worry about such things as the cost of rent, heat and light, the costs of employees' pensions and, now, the cost of paying people to stop work early. They also have long-standing commitments overseas that go beyond the well known obligations to subscribe to international laboratories but which entail collaboration with comparable research councils elsewhere (the Dutch for example) which cannot be quickly shuffled off. Then, more recently, it has been thrust on everybody's attention that universities are no longer able to fulfil their part of the bargain that people applying for research grants should be sufficiently well equipped to carry out the work contracted for.

Few doubt that if the research councils were free to start all over again, they would organize their affairs quite differently. They would avoid like the plague long-term commitments to people, or in respect of buildings. That they are so lumbered is partly the consequence of what now seem historical accidents but which are the legacies of past convictions that purpose-built laboratories or institutions were necessary for one vital purpose or another. Farmers are bothered by weeds among their crops? Then let there be a Weed Research Organization (now to be disbanded). Britain must keep up with the Joneses of Europe in high-energy physics? Then let there be not merely CERN but two substantial laboratories in Britain, in Oxfordshire and Cheshire (which have since been converted for other uses). Britain needs more clinical research? Then let there be a Clinical Research Centre (distinct, believe it or not, from the National Institute for Medical Research). The only comfort for the research councils is that other British public agencies, the Ministry of Defence and the Department of Industry for example, are also encumbered by such relics from the past, some of them much more expensive.

Encumbrances

If the start-again recipe would yield research councils free from these encumbrances, why should not some such goal be their objective for the years ahead? As the Agricultural and Food Research Council knows to its cost, and the Natural Environment Research Council only a little less vividly, that is the direction in which external pressures have been working in the past few years. The Medical Research Council's clumsy announcement last week of its decision to reduce spending within its own laboratories for the sake of being able to sustain spending on untied research is similarly a step in the right direction (see p.304). But the process has no gone nearly far enough. The research councils as a bunch now have no choice but to reduce as fast as they can their long-term commitments to people and in respect of buildings and fixed equipment, keeping only what can be held to be strictly important to the research enterprise. Until they move decisively in that direction, the councils must live with the problems they have made for themselves.

The machinery for making decisions about spending money on basic research would also come out differently if the whole system were being recast. The case of the Science and Engineering Research Council, the largest of them all and the chief source of funds for university-based research, is a good illustration of what has gone wrong. Each year the council, having fought the other



research councils for its allocation of public money, parcels out what it has to spend among a number of separate baronies responsible for research in different fields — astronomy, nuclear physics and so on. Democracy by committee being what it is, the result is that only marginal changes can be accomplished from one year to the next in the general pattern of spending.

But how else to deal fairly with a torrent of research grant applications when funds are short? The catch implicit in that reasonable question is that the members of the network of cheeseparing committees allow their sense of fair play to show as caution when the budget is tight, with the result that they back dull projects more often than they should. Sticking pins in a list of research grant applications would not suit either. On the start-again recipe, the research councils would instead follow different strategies, spending part of their funds with universities or university departments which had performed well in the past, sometimes inviting bids for research support in fields likely to have promise and conducting much more of their business in ways that would engage interested sections of the scientific community in decisions about what should be done. All these stratagems could of course give great offence, partly because they would constitute a departure from what has always been done, partly because they would seem less fair and, paradoxically, too open at the same time.

But is it not dangerous to acknowledge, when faced with pressure, obscurantism or even ignorance from above, that steps could be taken to improve the administration of the research enterprise? Only if the managers of the enterprise are not clear what they want to do with it. Now, there should be no doubt in anybody's mind — the tenor of British basic research needs to be strengthened, and effective steps need to be taken to see that the money being spent yields a substantial practical benefit. The British Government has given the research councils a licence to pursue the first goal, and keeps reminding everybody of the second.

The obvious difficulty is that none of the research councils nor their supervisory body, the Advisory Board for the Research Councils, is properly constituted for the pursuit of long-term objectives. Because of the inertia of the committees with which all are lumbered, clarity of purpose is rarely possible, but because both committee members and full-time officials play musical chairs with each other, continuity is equally elusive. (Sir James Gowans, secretary of the Medical Research Council, has nevertheless just been appointed for a second five-year term.) The result is almost the antithesis of leadership. New people arrive, seek to make their mark, start up some new programme which costs little at the beginning — and in the process lumber their successors with a range of expensive projects not fully understood. For a nation as obsessed as the British have been for three decades with the contemplation of the administration of research, it is astonishing that they have been so luckless at devising a strategy.

The remedy is simpler than it seems, but unfashionable. In the peculiar constitutional circumstances of Britain, the only means that has been found of combining policy with persistence is political. If there is a minister in charge of something, even the administration of bird sanctuaries, he will sooner or later be required to make a policy statement on the subject and that policy will be public policy until a successor makes a different public statement. Such a system could be operated on behalf of British basic science even as things are, for there is a minister (Sir Keith Joseph) with a bureaucracy to support him (the Department of Education and Science). But quite apart from the minister's obvious disaffection from the task, the department is unsuitable. The science budget has always been peripheral to its main business, the continual reorganization of schools and school systems. Especially because the need for a seemingly coordination of what is done on behalf of civil science with what other government departments do in competition has been made more urgent by the latest crisis, the time has come to revive the arrangement briefly (and successfully) relied on in the early 1960s — a part-time minister, a small bureaucracy and some open discussion of what should be done. □

What limits for coal?

The neglect of coal is proof that coalmining is a dirty job whose cost makes the product dear.

WHEN people used to talk, in the early 1970s, of the physical limits that would eventually halt economic growth, perhaps catastrophically, it was customary to cite materials such as tungsten or lead as examples of resources that were undeniably finite. The gloomy prophecies of those now-distant times are chiefly memorable for their failure to predict that an insubstantial circumstance such as a paper increase of the price of oil would do the same trick, and much more quickly, but they were also a rich source of prognostication about the future supply of energy in advanced societies. Briefly, the argument used to go, petroleum (the chief fuel of the 1960s) was a diminishing resource destined to become scarce and expensive in the then foreseeable future, perhaps in fifty years, certainly within a century. And since there were objections to the widespread use of nuclear power apart from the physical shortage of uranium, and because of the world's shocking neglect of resources such as solar energy (now called "renewables"), there would be an uneasy period in the future when it would be necessary for people to rely on coal as the chief fuel. Thankfully, most prophets would remark that there were at least ample reserves of coal still to be extracted from the Earth's crust.

So what is to be made of the present trouble in the British coal industry, where annual coal production has fallen steadily since the 1970s, where the labour force has declined even faster (because productivity has increased) and where a third of the labour force was on strike last week in protest at the industry's plan to reduce output still further, by 4 million tonnes from last year's output of 98 million tonnes? There is nothing wrong with the simple arithmetic of the old forecasts. Britain, which has been winning more than 100 million tonnes of petroleum from oil fields in the North Sea for more than the past decade, will soon see this windfall resource begin to tail off; production from the fields now being worked will be declining in aggregate before the end of the 1980s. And while there is no prospect that nuclear power will take up the slack, reasonable people might think that coal would be in demand again. Yet the opposite is true.

The explanation of the unexpected neglect of coal in the past decade is primarily the recession caused originally by the increased price of oil but which now also owes much to the permanent improvement of fuel efficiency to industry, commerce and in domestic usage. Those who constantly cry for what they call a fuel policy — by which they mean a set of explicit targets specifying the quantities of different kinds of fuels that will be used at different times in the future — should reflect that there has hardly ever been a better example of how usage will find a natural balance under the influence of market forces. The high cost of fuel has done more for fuel efficiency than any exhortation.

In Britain, however, there has been another uncomfortable economic influence at work — the unwillingness of coalminers to allow their employers to close down coal-pits operating at a loss. The result is that the coal industry will this year make a loss of more than £200 million even after counting the cost of public subsidies for stockpiling unsaleable coal, for not importing cheaper coal from elsewhere and for the social costs of uneconomic operations. The miners' chief argument, that it is somehow morally wrong to walk away from coal deposits that could be worked but only at high cost, resembles the old argument about the limits to growth in that it ignores economic costs. In present circumstances, the argument threatens to bring about a further and needless reduction of output. But for the long run, the dispute has a more sinister meaning, for it implies what most people will acknowledge to be true — that coalmining is such an unwanted and relatively ill-rewarded occupation that the cost of getting people to do it will always increase to the point at which coal is as expensive as any other fuel on the market. So the huge reserves of coal may be less useful than they used to seem. □

High-school education

Damning accounting office report condemned

Washington

LEGISLATION committing hundreds of millions of dollars to new programmes designed to improve the knowledge of school mathematics and science teachers may have little impact on the declining performance of US schoolchildren. That is the unexpected conclusion of a heretical study just published by the congressional General Accounting Office (GAO). The study, which contradicts the recommendations of almost all the recent reports on the plight of US high schools, has provoked sharp criticism from the National Science Foundation (NSF), the Department of Education and the White House Office of Science and Technology Policy (OSTP).

Dr James Ling, assistant director of OSTP, has written to GAO complaining that the report is clearly biased against upgrading the quality of existing mathematics and science teachers and that it "suppresses" evidence that teacher retraining programmes have indeed improved the performance of students. Dr Edward Knapp, director of NSF, says the report is "unfortunate and counterproductive". The Department of Education says it disagrees with most of the report's major conclusions.

This angry reception can be traced in part to unfortunate timing. Concern about the poor state of mathematics and science education in the United States — reflected by a 20-year decline in scores on the Scholastic Aptitude Test — reached a peak last year. A national commission on education set up by the Reagan Administration said standards had declined so much that the nation was "at risk". And a comprehensive report by the National Science Board called for a \$1,000 million a year programme to retrain mathematic and science teachers and establish 2,000 "model" schools where innovative approaches to mathematics and science teaching could be encouraged.

These and other reports have stimulated a wave of educational reforms at state level. But there is precious little to show for them nationally except for two bills which, supporters glumly concede, stand little chance of becoming law before the presidential election at the end of the year. In March 1983, the House of Representatives approved an Emergency Mathematics and Science Education and Jobs Act. In the Senate, an Education for Economic Security Act was reported out of committee last May. Both bills would pump an additional \$400 million a year (enough to spend nearly \$2,000 on every secondary mathematics and science teacher) into mathematics and science

education.

Neither bill has been able to generate enough support to move much further. What support they do have could be seriously damaged if the GAO report is given wide credence. Central to both bills are proposals for teachers' institutes and other forms of in-service training designed to refresh the knowledge of teachers already employed to specialize in mathematics and science. Yet the GAO study, after reviewing attempts to do the same thing in the past, says there is a good chance that such measures will have no effect on the performance of the schools.

The report takes as its starting point the summer institutes organized by NSF between the mid-1950s and the mid-1970s. Participation was widespread. A 1971 survey found that 51 per cent of secondary science teachers had attended one or more institutes, established to "increase the effectiveness of teachers by broadening and updating their scientific background". The scheme ended in 1975 when the shortage of scientists became a surplus and NSF's curriculum development programmes ran into political hot water over a controversial social science course. But how effective, meanwhile, had the institutes been?

According to GAO, only one useful research project ever asked that question and its results were ambivalent. High-school students of teachers who had attended the institutes did better but junior high-school students did not. So GAO tried to answer the question a different way.

Mathematical and scientific knowledge, not teaching skills, dominated the institute courses. Do teachers who know more about their subject actually get better results from their students? The intriguing evidence from recent research is that they do not.

This finding, GAO argues, has important implications for public policy. Spending heavily to improve the knowledge of existing mathematics and science teachers, for example, may not be as effective as converting good teachers from other subject areas. And it may be better to concentrate, in retraining courses, on general pedagogic skills than on substantive knowledge of a discipline.

Both conclusions run counter to conventional wisdom. Its critics say the GAO report jumps to conclusions that cannot be supported by the slender research findings it reviews. Even so, the report is a disturbing warning that the problems of quality, if not quantity, of mathematics and science teaching may not be as tractable as supporters of expensive legislation believe.

In Illinois, for example, five laboratories belonging to the Illinois Research Corridor have banded together to offer summer jobs to superior science and mathematics teachers from their local schools. The five — Amoco Research Center, Argonne, Bell Labs, the Nalco Chemical Company and Fermilab — say the programme was designed to be interesting and relevant but contains no formal attempt to improve specific teaching skills.

Both the teachers and the laboratories which hired them claim, on the basis of the first summer, that the scheme has been outstandingly successful. More laboratories intend to participate this year. The teachers, now back in their schools, report renewed interest in teaching and increased confidence in their own abilities.

Peter David

Bug takes to the ski slopes

It seems that *Pseudomonas syringae*, the obscure Gram-negative bacterium that was recently catapulted to world fame when plans to conduct field trials with a genetically engineered version were blocked in the United States, may yet play a larger part in the affairs of men. Last October, plans for a field trial that entailed the release of genetically engineered *P. syringae* lacking the bacterium's ice nucleation protein had to be postponed after objections from environmental groups. But researchers at Advanced Genetic Sciences Inc, at Oakland, California, hope to use other organisms expressing the same protein to gain a lead in the lucrative market for artificial snow used on ski slopes.

The use of nucleator makes snow-making much easier than it would otherwise be, as it largely prevents the water from becoming super-cooled. Water is simply sprayed

through a fine nozzle onto a fan, and the expansion-induced cooling produces snow. But there are problems with existing nucleation agents, and the *P. syringae* protein fits the bill perfectly.

Dr Trevor Suslow, who is working on the project, says the company is ready to scale up to full production with a system using whole freeze-dried *P. syringae*. Although the bacterium is harmless, it will be killed — probably by gamma irradiation — before use, if only to reassure skiers who lack a knowledge of basic bacteriology.

With this system, ice forms at about -3°C . But when the regulatory hurdles standing in the way of releasing genetically engineered organisms are overcome, the company hopes to use other bacteria to produce the protein. An even more efficient and inexpensive ice nucleator may then become possible.

Tim Beardsley

UK research councils

Bolt hits medical research

THE directors of the research units of the British Medical Council (MRC) have been abruptly informed that their budgets for the financial year beginning in just ten days on 1 April will be much tighter than anticipated. Several directors describe themselves as "shell-shocked" by the news, given to them at a meeting last week with the council's secretary, Sir James Gowans. The amount available for capital equipment such as instruments will be cut drastically to around £1 million, and operating budgets will be cut by approximately 20 per cent. Salaries will not be affected, nor will the number of project grants to universities, but posts may be left vacant to make savings.

Substantial cuts in MRC's budget had been expected in the financial year 1985-86, to allow for the "restructuring" of the Agricultural and Food Research Council (AFRC) and the Natural Environment Research Council (NERC), both of which are planning staff reductions that will entail redundancy payments. This year's cuts, however, are apparently unrelated, and reflect the council's decision to spare research in the universities and even to compensate them for cuts to their recurrent grant paid through the University Grants Committee. The total budget for MRC is to be £117.2 million, a cut in real terms from last year's budget of £113.2 million.

MRC unit directors last week said that the 16 per cent cash cut to their recurrent "consumables" grant, which will save some £3 million, does not tell the whole story. Unlike previous years, there will be no automatic allowances for inflation, so the cut in real terms is about 20 per cent. Furthermore, the costs of laboratory sup-

plies are running well ahead of the inflation rate, as are fixed costs such as property taxes. Directors complain that they are committed to meet the cost of orders already placed. The steep drop in the level of the capital equipment grant, from £5.8 million in 1982-83 to £1 million next year, means that equipment purchases will be restricted to a few large items and those necessary for safety. MRC told directors last week that an extra allowance for capital equipment may be paid in the autumn.

It is not yet known how the individual MRC units will be affected. At last week's meeting, Sir James said that some units will be spared any recurrent grant cuts whatsoever, while others may be cut by more than 20 per cent. Units that have new programmes or unusual fixed costs — such as the Clinical Research Centre — will probably fare better than others. The council has had to meet the costs of establishing four new units, and their programmes are considered inviolate. And it has, like others, been hit hard by the obligation imposed on it to increase its contributions to employees' pension funds, costing £600,000 next year. MRC also lost about £500,000 as a consequence of a rounding-off exercise in the allocation of funds between research councils.

SERC will be making its financial allocations next week. Although the council will in 1985-86 be losing funds to support restructuring in AFRC and NERC, an official denies that SERC plans to make drastic cuts along the lines of those at MRC. SERC's budget increased this year by 10 per cent, to £278.8 million.

Tim Beardsley & Stephen Budiansky

Laboratory animals

Primate supply shrinks

THE increasing difficulty of importing research primates is forcing European countries to establish breeding colonies of commonly used species. Malaysia last month became the latest of a series of countries to ban the export of primates for research, citing as its principal reason the discovery that monkeys exported to the United States were being used for biochemical and nuclear weapons research. There were also worries over the size of the remaining populations in the wild.

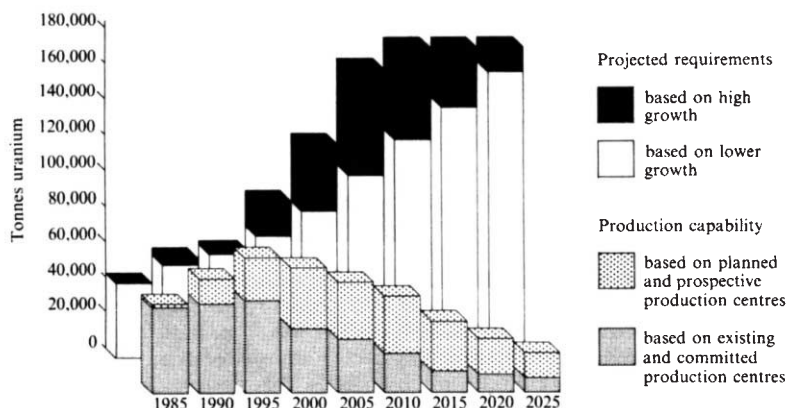
Military research accounts for about 15 per cent of the 20,000 primates used each year in the United States and about 12 per cent of the 3,000 used in Britain: the great majority are used for testing vaccines. But 12 countries have so far acted to prevent the export of their primates, for various reasons, and if Indonesia and the Philippines were to follow suit, prices would rise drastically.

Attempts to establish breeding colonies in the countries of use have not often been economically viable, producing (for example) rhesus monkeys at a cost of around £1,000 — two to three times the cost of an animal caught in the wild. But the governments of West Germany and Switzerland have been persuaded of the need to establish national breeding colonies for primates, and in Britain the Medical Research Council is encouraging a switch from Old World to New World monkeys, which it breeds itself.

Malaysia, which exported between 2,000 and 3,000 macaques a year, was not a major supplier and its ban is not expected to lead to acute shortages. But the decision by India to ban the export of rhesus monkeys in 1978 led to a switch towards the use of macaques, and future political moves could lead to further changes in the species used. While the wild populations of macaques in Indonesia and the Philippines are not yet under serious threat, some in the primate business are realizing that the supply will not last for ever.

Mr Keith Hobbs of Intersimian Ltd is one such. Hobbs says he left the Medical Research Council in frustration that the council would not heed his advice to establish a breeding colony to meet Britain's primate needs. His company, through an organization called SICON-BREC, is now building up a breeding colony of macaques in the Philippines. Hobbs stresses that the high rates of loss in transport that were common in the past have now been much improved. His aim is to build up a breeding colony of around 1,000 females — which would go a long way towards meeting Britain's needs. His price for a fully conditioned *Macaca fascicularis*: just £200. Tim Beardsley

Oil crisis . . . uranium crisis?



If worldwide production of uranium continues at its projected levels it could severely restrict future nuclear programmes, despite the fact that stockpiles are growing. This is the conclusion of a report released recently by the Nuclear Energy Agency of the Organization for Economic Cooperation and Development (OECD) and the International Atomic Energy Agency (*Uranium: Resources, Production and Demand*). Stockpiles should hold until the end of the 1980s, but then demand will far outstrip production unless new uranium production programmes are started. The period since 1980 has seen a fall in production by the United States and increased production from Australia, France, Brazil, South Africa and Namibia.

UK research

Prospect of squalls ahead

THE heads of Britain's five scientific research councils have temporarily shelved their differences to make an unprecedented plea to the government for a substantial increase in next year's science budget. They say that valuable new scientific opportunities are being lost because the government's "level funding" policy takes no account of the rising costs of scientific instrumentation and raw materials.

The plea is made in a paper drawn up by the research council heads under the chairmanship of Professor John Kingman of the Science and Engineering Research Council. The paper was due to be discussed this week at a meeting of the Advisory Board for the Research Councils as the first stage of its annual survey of scientific opportunities, when it will argue the case for an increased science budget in order to give the Secretary of State for Education and Science, Sir Keith Joseph, ammunition for battles within the government over public expenditure priorities.

Last year, the research council heads each made separate bids for support of important new developments in their own areas. That was entirely unsuccessful: the £27 million extra they requested was flatly turned down, and the only increase in real terms was earmarked to meet the rising cost of international subscriptions. As a consequence they have this year sought strength in numbers by throwing their lots in together.

The government's case is that the size of the science budget (£550 million in the coming financial year) has been roughly maintained in real terms, and that Britain must retain its high standing by making savings in less important areas and by increasing efficiency. But the research council heads emphasize that the extra five per cent on top of the existing science budget that would be needed for new areas to be exploited cannot be found by increasing efficiency or by diverting funds from other areas.

Sir Hermann Bondi, chairman of the Natural Environment Research Council, says that while he is willing to divert "a few millions" into priority areas, it is impossible to achieve savings to meet the costs of space-based research, which are of "a different order of magnitude" from conventional scientific activity. This "sophistication factor" applies also to Earth-based instruments as they become computerized, says Sir Hermann. He stresses that science is an "internationally competitive business" and that the United States, France and Japan are all sharply increasing the levels of financial support for science. Britain's continued participation in the extension of the International Deep Sea Drilling Project is still in doubt due to pressure on funds.

Both the Science and Engineering

Research Council and the Natural Environment Research Council are worried about the steep drop in the proportion of project grant applications deemed to be of "alpha" merit that they are turning down on cost grounds. In engineering, for example, some 40 per cent of alpha-grade applications are now being refused. And there is increasing concern that Britain is relying for space-based work on the charity of other nations. The Science and Engineering Research Council argues that in order to maintain its international standing, Britain must at least occasionally lead a satellite programme. But at present there appears to be little chance of that happening.

The Medical Research Council is also hard hit by increasing costs (see p.304). Plastics and restriction enzymes are cited as examples of essentials that are increasing in cost faster than inflation. Pay increases to

staff based in universities and to civil servants have exceeded the allowances made for them and are outside the control of the individual councils. The rising cost of pension contributions (which have to be paid for out of the research councils' budgets) is another obligation which reduces the value of the science budget and which is likely to increase. The council would be one of four that would collaborate in a £4 million initiative on biotechnology if funds were available.

The Advisory Board for the Research Councils will include the research heads' document in its advice to Sir Keith next month. But once the likely size of the 1985-86 science budget becomes known, the temporary alliance may come to an abrupt end as the councils fight to decide how the cake will be divided between them. Last year's advice on the division of the science budget is due to be published, belatedly, this week. Participants in the exercise are preparing themselves for some close questioning on their priorities.

Tim Beardsley

Polish academics

Nuclear institute still disfavoured

POLISH scientists who worked at the former Institute of Nuclear Physics (IBJ) at Swierk outside Warsaw are said to have rejected a bargain that could have led to the reinstatement of a number of scientists dismissed during the martial law period. During the winter of 1982-83, the institute was closed down and divided into three.

Several independent reports reached *Nature* at that time that at least 40 researchers who had held tenured positions lost their jobs and that those affected were for the most part former activists of Solidarity.

According to the underground Solidarity press, the Party authorities intimidated at the end of last year that they were prepared to "reconsider the question" if the Swierk scientists would undertake certain "political initiatives":

- To draft and sign a statement condemning the siting of Cruise and Pershing missiles in Western Europe. (This must be signed by not fewer than 40 persons holding the degree of *Doctor Habilitatus*.)
- To "deny in *Nature*" (presumably by means of a letter to the correspondence page) the article published in 1983.
- To exert influence to counteract "terrorist actions" on the premises of the former IBJ.

DESMOND on line

THE development of computer acronyms will pass another milestone this week when the Open University in the United Kingdom unveils its new educational machine DESMOND — the Digital Electronic System Made Of Nifty Devices. □

- To write, for example in *Polityka*, the influential weekly paper, an article "to be based on material supplied by the Central Committee's Science Department" appraising "in the correct manner" those physicists who have gone on official trips abroad and then stayed in the West.



The "terrorist actions" referred to in this list refer, presumably, to the strike at Swierk in protest against the imposition of martial law in December 1981, and the possibility of further protests by Solidarity supporters at the institute.

According to the underground press, the Swierk scientists have refused to comply with the proposed "initiatives" on the grounds that the authorities have no right to make such demands.

Nature would, of course, be delighted to reassure its readers that all is well at Swierk — as soon as clear evidence is available that this is indeed so.

Vera Rich

Soviet academics

No room at the top?

THE Soviet Union, with more than a million people of working age with higher education, is having difficulty finding jobs for them all. The rapid expansion of higher education during the past three decades to meet the challenge of the "scientific and technological revolution" has proved, it appears, too successful. Some universities and institutes seem now to be attempting to obstruct researchers hoping to submit a dissertation for a higher degree on the grounds that no posts are available for them.

The problem is exacerbated by the Soviet two-tier system of higher degrees and by the rigid separation of available posts into those for "junior" and "senior" scientific researchers. It is still fairly easy for a Candidate of Sciences (roughly equivalent to a PhD) to get a position as a "junior scientific worker". The trouble comes when he or she wants to proceed to the degree of Doctor of Science. If there are no senior posts available at an institute or university, the argument goes, why should a junior scientific worker or assistant lecturer want to take a doctorate? A doctor cannot be employed in a junior post, so anyone who insists on submitting a doctoral dissertation must do it elsewhere.

During the past few years, the issue of higher degrees and the jobs appropriate to them has been taken up on a number of occasions by the weekly *Literaturnaya Gazeta*, which published in 1976 a proposal that the difficulty could be avoided by dropping the word "junior". In 1982, it questioned the relevance of the candidate degree (introduced as a temporary expe-

dient in the 1930s to satisfy an urgent need for trained personnel). Last week, it launched yet another "debate" on the subject, with a letter by Academician V. L. Ginzburg, of the Landau Institute of Theoretical Physics, entitled "Is it necessary to block those presenting dissertations?" — a question which he answers with a firm negative.

The publication of scientific articles and the defence of dissertations are, in Ginzburg's opinion, the "basic and often unique form" in which a scientist can demonstrate his or her results. Placing impediments in the way of these activities, he says, imposes on victims a "moral trauma" and takes away their legal rights. Jobs, however, are a quite different matter, and if a new doctor of science wants to stay on in his old (junior) job, that is his business. (In Soviet conditions, in which the preparation of a doctoral dissertation can take up 10 to 15 years of working life, it is hardly surprising that Ginzburg regards it as a scientist's "basic" activity.)

Ginzburg argues that the defence of a dissertation for a higher degree should be free of administrative restraints — though subject, of course, to those imposed by the Higher Attestation Committee (which, incidentally, cover character and political rectitude as well as purely academic considerations). He does not mention, however, the main practical objection to Doctors of Science in "junior" posts — the Soviet academic salary structure which calculates salaries not on the basis of the post held but on the qualifications of the incumbent.

Vera Rich

Insurance wanted on Greenland explorers

EXPEDITIONS planning to visit Greenland are likely to find it more difficult to raise adequate support from this year. The Commission for Scientific Research in Greenland is insisting on much larger guarantees against the cost of rescue operations — twice as much by some accounts — and visitors will in future have to liaise more closely with the authorities and stick to an agreed route. The number of private scientific expeditions may fall as a result.

The commission, based in Copenhagen, says the tight new conditions have been imposed because of several fatal accidents last year. Although all the fatalities occurred on sporting rather than scientific expeditions, the commission is taking no chances. All agree that the cost of rescue operations in Greenland is enormous: helicopters, usually the only means of rapid transport, start at \$1,500 per hour. The commission and the Ministry for Greenland in Copenhagen have been frustrated at the number of expeditions setting off into remote Greenland with what is seen as

wholly inadequate equipment and with little or no experience of Arctic travel.

There may also be a political element to the decision. Although still formally a part of Denmark, Greenland now has effective home rule, and — as is shown by its recent decision to leave the European Economic Community — is willing to go its own way when the occasion arises. As Denmark already subsidizes the country to a large extent, attitudes may be hardening.

The commission is adamant that *bona fide* scientific expeditions will still be welcome in Greenland. Its aim seems to be to eliminate many of the badly organized sporting expeditions which have been increasing in recent years. Well-planned expeditions, it says, will not be imperilled.

Others are less sure. There were 90 scientific expeditions to Greenland last year, of which 64 were not Danish. The cost of insurance — proof of which must be shown to the commission — may be a significant part of the cost of an expedition, particularly to a remote area.

Tim Beardsley

US research freedom

Suit against farm machines

Los Angeles

THE freedom of a university to choose its own research objectives went on trial in California last week.

In an unprecedented court case, a group of farm workers challenged the right of the University of California to use public research funds to develop labour-saving machines that, according to the farm workers, primarily benefit owners of large farms.

The 19 workers say mechanized devices developed by the university's agricultural extension division have eliminated thousands of farm jobs, promoted the growth of large land holdings and prevented workers from starting small farms of their own.

The university denies the charges and says that no more than 3–4 per cent of all its agricultural research has ever gone to developing the labour-saving machines. It also says that much of its research has directly benefited both small farmers and farm workers.

The case is being tried in Alameda county, across the bay from San Francisco and just south of the Berkeley campus. Plaintiffs are represented by California Rural Legal Assistance, a federally-subsidized public interest law firm.

The workers are seeking a court order that would limit the influence of private industry in helping to set research priorities for university scientists and to require that researchers prepare "social consequence statements" before taking on projects.

The consequences of stopping research because what might be learned could some day be harmful "is staggering", said university lawyer Gary Morrison. "Professors should be allowed to inquire into areas of academic merit, where they believe that new knowledge should be uncovered, without judicial interference and without having to consider some perceived downstream, social or economic impact that might occur."

The judge in the case agreed that it is difficult to determine in advance what the economic and social impacts of new knowledge will be. But he said he would allow the plaintiffs to determine if the university's procedures and policies favour private over public interests.

University lawyers have likened the farmworkers' challenge to last century's Luddite revolt to stop the use of labour-saving machines. But Ralph Abascal, representing the California workers, said the comparison is unfair and "that Congress intended the system of public-supported agricultural research to help the little person, the person most in need, not mechanization research for large industries".

Sandra Blakeslee

UK geophysics research

More support for solid-Earth?

THE community of solid-Earth geophysicists in the United Kingdom is inadequately represented and supported by the Natural Environment Research Council (NERC). That at least is the message from a recent survey of university Earth scientists carried out under the auspices of the Royal Astronomical Society. And, while NERC rejects some of the charges levelled at the way it assesses studentship and research proposals, there is no argument that the level of geophysical activity in the United Kingdom represents a significantly small proportion of efforts in the Earth sciences as a whole.

The charges implicitly levelled at NERC arise from a survey carried out by Professor K. Creer, head of the department of geophysics at the University of Edinburgh; those who responded included geophysicists and geologists from many British universities. The areas of concern include electromagnetics, geomagnetism, gravity and geodesy, palaeomagnetism, seismology and planetology.

One worry, according to Professor Creer's report on the survey, concerns the way in which NERC committees operate. The training awards committee for the Earth sciences, consisting of 17 scientists, does not use peer review in reaching its decisions — because according to NERC, the resources are not available to cope with such an assessment of 600 or so applications each year, of which about 130 are successful. Moreover, NERC's view is that anyone assessing an application within the United Kingdom would be a competitor, and that it would be impractical to have the applications assessed overseas.

The survey highlights the difference in classification between geophysicists with geological backgrounds on the one hand and those who apply expertise in physics and mathematics to geophysical problems on the other. If the latter category is accepted as the "pure" definition of a geophysicist then, over the past decade, fewer than 10 per cent of the members of training awards and research grants committees have been geophysicists.

NERC's response is that it is keen to ensure that committee members are not seen as representatives of particular lobbies; rather, they should be capable of judging across all the fields concerned, with the help of peer review.

One feature of NERC policy highlighted by the survey is a concentration on crustal studies. Most geophysicists who responded work in this area and were satisfied with NERC's approach, except to question the relatively small amount of NERC's resources put into university research rather than other NERC activities (especially its own research institutes). The survey says, however, that such areas as electromagnetic induction, physical prop-

erties of rocks and palaeomagnetism with geomagnetic objects are experiencing undue difficulties — a point that NERC simply rejects. One significant area of crustal research is the deep seismic crustal profiling project (affectionately known as BIRPS). The survey report points out that, as this involves extensive outside contracts in its operation and crustal geological interpretation, it should hardly be counted as real geophysics, and that, moreover, it has prejudiced support for geophysics by its sheer financial requirements. On the latter point, NERC says it was encouraged by ABRC to concentrate on certain areas likely to be particularly productive — an approach amply justified, it says, by the results of the deep seismic reflection projects.

A major issue reflected in the survey concerns the coordination of support for geophysics between NERC and the Science and Engineering Research Council

(SERC), an issue that seems to occupy the minds of UK geophysicists of all persuasions. Indeed, the Royal Society working party looking into support for geophysics as a whole (including atmospheric and magnetospheric science as well as the Earth sciences) was stimulated to do so partly by the feeling in the communities that the atmospheric/space geophysicists are insufficiently strong in the competition for SERC funds compared with the astronomers, while some solid-Earth geophysicists feel unduly dominated by the geologists on the NERC committees.

That the research councils are aware of the problem is reflected in a recent meeting between Sir Hermann Bondi and Professor John Kingman, chairmen of NERC and SERC respectively. They concluded that the present distribution of responsibility between the councils is appropriate but that more attention needs to be given to the borders of responsibility. To this end, Dr John Mather (NERC) and Dr Barry Martin (SERC) have been given responsibility for ensuring adequate cooperation.

Philip Campbell

UK science parks

A growing phenomenon

THE year 1984 looks set to be a bumper one for science parks in Britain. Although opinions differ on what they should do and even on what they are, everyone is convinced they are a good thing and should be encouraged. So far this year, the University of Birmingham has officially launched its grandly-styled "Institute of Research and Development", Barclay's Bank has opened a "Venture Centre" at the plain old science park at the University of Warwick, while the University of Stirling hopes this week to reach agreement with local authorities over its long-delayed proposed science park (not to be confused with the High Technology Area, over which the university relinquished control by selling it to Wang computers). Even the polytechnics are getting in on the act: the Polytechnic of the South Bank this week plans to open its "Technopark" on a derelict site in South London.

Connoisseurs of the species would probably object to the last-mentioned being described as a science park, since it is neither exclusively scientific nor park-like. There is a distinction to be drawn between science parks opened in the north of England, which have been seized upon by local authorities as a regional development tool (as well as by universities as a source of extra income), and the more spontaneous southern type. And investors are becoming more wary of the idea of science parks, some of which seem hard to distinguish from up-market office accommodation.

As these enterprises were seen chiefly as a means of attracting new industry to depressed areas it may, however, be unfair to judge them too harshly. The true science

park has to have a low building density and be in an attractive environment. One of the main lessons of the British experience so far is that the sort of people with ideas that will regenerate industry are rather fussy about where they want to live. Cultural amenities are as important as close proximity to a university, it seems: one recent study found that technology-minded companies are as likely to forge links with universities elsewhere in the country as with the one on their doorstep.

The investor, even if he has lost a little of his initial enthusiasm, is not yet totally disillusioned with science parks. Although seen by some universities as a panacea for all ills, few are likely to become mature investments within 15 years. The science park at Cambridge, the oldest in the business, is now said to be making a good commercial return on the investment of those who backed the project, although it has now relaxed its original rule against any manufacturing industry on the site. Heriot-Watt University in Edinburgh has a research park, which it insists will never be sullied with manufacturing industry.

At Birmingham, the university has set up a management company to head its institute. It has been luckier than most in having "loads of money" to spend on the project, in the words of John Samuels, Birmingham's professor of business finance. Although it does not intend to follow others and tempt industries to move in by offering low rents, Professor Samuels says that several companies who want to collaborate with university academics are prepared to take the plunge.

Tim Beardsley

Three Mile Island

Still a problem after five years

Washington

THE fifth anniversary of the accident at Three Mile Island has brought little tranquillity to the reactor site at Harrisburg, Pennsylvania. Efforts to clean up the damaged Unit-2 reactor are slowing down because the General Public Utilities (GPU) Nuclear Corporation is running short of money. Meanwhile, the site continues to provide a focus for bitter arguments about the safety of nuclear power in the United States as well as the integrity of both the industry and the Nuclear Regulatory Commission (NRC).

GPU technicians have not yet begun to remove the radioactive debris from the reactor vessel, although the company planned last week to extract a second batch of samples from the fuel core. Later in the year, 60 studs on the head of the reactor vessel will be loosened to see whether any have been so damaged by corrosion that they cannot be removed when the time comes to lift off the head.

Progress on the clean-up is, however, being hobbled by GPU's financial difficulties, caused principally by the industry's sluggish response to appeals for financial help. Some \$400 million has already been spent on the clean-up, but GPU says more than \$1,000 million is needed. In July 1981, Pennsylvania's Governor Thornburgh proposed a plan under which GPU itself would pay a quarter of the sum, and nuclear utilities a fifth. Another fifth would come from federal matching grants, and the balance from insurance payments and the governments of Pennsylvania and New Jersey.

So far the industry has held back. The Edison Electric Institute reports pledges from industry of \$71 million, but will not release those funds until at least \$100 million has been committed. However, utilities may soon prove more forthcoming as the result of an Internal Revenue Service ruling last December that contributions to the clean-up may be deducted as an ordinary business expense.

GPU's financial predicament may also be relieved if it can persuade NRC to let it restart the undamaged Unit-1 reactor at Three Mile Island. The commission hopes to make a decision by the end of June, and there are signs that it will give GPU a clean bill of health. Most encouraging for GPU was a decision in January by NRC to set aside questions of management integrity when it comes to vote on the restart. But the issue of integrity has recently taken on new importance in the wake of a grand jury indictment accusing the Metropolitan Edison Company, which operated Three Mile Island at the time of the accident, of deliberately falsifying coolant leak rate tests for months before the accident.

Had the leak rates been reported in accordance with NRC regulations, the

plant would have been shut down and the problem identified. The accident occurred when the plant operators ignored control room indicators of high temperature in the system. They did not discover until after the accident that a relief valve was open, allowing the core temperature to rise and hundreds of gallons of coolant to escape. In a plea-bargain agreement last month, Metropolitan Edison admitted its guilt. It is to pay a \$45,000 fine and prosecution costs. In addition, the company must establish a \$1 million account to help the Pennsylvania Emergency Management Agency to formulate an emergency plan for a 20-mile zone around the plant.

Even before the grand jury indictment, GPU had been plagued by allegations that its management of the rescue operation had been incompetent and, on occasions, unethical. Its most implacable critic is NRC commissioner Victor Gilinsky, who insisted last year that the senior management of the company would have to be replaced before he would support a vote to restart Unit-1. GPU, Gilinsky complained, had a "narrow and grudging conception of its public responsibilities", cut corners on safety and held back information from the public authorities.

GPU's relations with the commission have certainly been unhappy since the accident. NRC investigations have been held to find out whether GPU employees cheated during training programmes, whether engineers who expressed concern about safety issues were victimized by management and whether internal accident reports were doctored before they were submitted to NRC. In a report last September, the commission said allegations that GPU had violated safety procedures were true. An investigation of the victimization charges is still incomplete.

So far, the corporation has denied all the charges against it. While conceding that some technical violations of safety procedures had taken place, a GPU-sponsored investigation said that they had been accidental technical errors with no direct impact on safety at the site. The corporation also persuaded retired admiral Hyman Rickover, former head of nuclear operations for the United States Navy, to conduct an independent investigation. In September he declared GPU fit to operate the plant. But the corporation has also bowed quietly to NRC criticism. In November the company's president and senior deputies were reassigned.

These moves may well persuade NRC to permit a restart at Unit-1 but they are unlikely to still public controversy. The Union of Concerned Scientists claimed last month that GPU had not yet introduced the technical modifications needed to ensure the safety of Unit-1. **Peter David**

US accelerators

Super-collider costing

Los Angeles

SEQUESTERED in special offices at Lawrence Berkeley Laboratory at the University of California, a group of 20 high-energy physicists is working round the clock to establish "credible" cost estimates for the United States' proposed superconducting super collider (SSC). The accelerator, which would collide two proton beams at energies up to 40 trillion electron volts, would be the largest and most powerful in the world.

The group, representing seven institutions and leading competitors who want to design SSC three different ways, expects in three months to produce a 200-page report for the Department of Energy (DoE) detailing how much competing designs will cost.

An important goal of the SSC reference design study, according to Jay Marx of the Berkeley laboratory, is to "do an honest job from bottom up and see what the numbers come out". It will not select a winning design but will compare the favourite approaches on a systematic basis.

"A lot of numbers are floating around this town", said Bill Wallenmeyer, head of DoE's division of high-energy physics in Washington, DC. With some estimates topping \$8,000 million, he said, the government needs "a better handle on costs".

Impetus for the cost study came from the directors of US physics laboratories. Meeting at Cornell University last September, they realized that while many designs were being discussed, costs were pure guesswork. Secretary of Energy Donald Hodel said he wanted to move on SSC by the summer of 1984. As a result, the study group began formal meetings on 1 February and will deliver its report to DoE on 30 April.

According to DoE, the \$2-million study is top priority. "We're expecting (it) to give us 30 to 50 per cent accurate cost estimates", Dr Wallenmeyer said. To meet a 1988 construction start, he said, the government must have credible cost estimates this summer.

In response to fears that a reference design study might "lock in" one approach too early, the group is comparing three designs. Trade-offs in technology, such as type and strength of magnets, determine costs. Some designs favour an accelerator with a 150-mile circumference while others, backed by those without access to cheap land, have a 30-mile circumference.

The mood at Berkeley is very upbeat, according to Dr Marx. "There are no winners or losers in this exercise and no sense of our design versus their design." US physicists are primarily interested in seeing that SSC is built, he said.

Sandra Blakeslee

Britain's binary higher education

SIR — Few would disagree with your view (*Nature* 2 February, p.400) that the United Kingdom's binary system of higher education is divisive, but I feel that your leading article contained some errors.

The local authority, or maintained, sector of higher education consists of many more institutions than the thirty (not twenty-six) polytechnics in England and Wales. These include the colleges and institutes of higher education and the Scottish central and local authority institutions.

Polytechnics were not designated as degree-giving institutions. They, in common with other maintained institutions, had in their constituent colleges offered degree courses for many years, usually validated by the University of London. Today, polytechnics are proud of the range of courses they offer leading to higher diplomas, professional qualifications and degrees — the degrees being validated by the Council for National Academic Awards.

It is misleading to suggest that the polytechnics have attempted to ape the universities. Their range and diversity of work, the variety of modes of study, their vocational emphasis, links with industry, commerce and the professions and ties with the regional and local community combined with their national roles fully justify their claim that they have been highly successful in achieving the objectives set out in the 1966 White Paper which included the establishment of "a strong and distinctive sector which is complementary to the universities".

Turning to the proposed Polytechnics Central Admissions System (PCAS), it must be placed on record that the initiative came from the Committee of Directors of Polytechnics (CDP). It is CDP which has successfully negotiated with the University Central Council on Admissions (UCCA), not the Department of Education and Science (DES), and CDP which has successfully negotiated with DES the offer of a grant towards the cost of setting up PCAS. What should also be placed on record is CDP's appreciation of the positive response of UCCA and of DES. Using your own wording, it is a nonsense to say that PCAS will run for a trial period of two years. It is anticipated that there will be two financial years in which expenditure will be incurred in setting up the system before any income received from participating institutions and applicants. Once the system has been established, it is expected to be self-financing.

In suggesting that confusion may arise because applicants using UCCA and PCAS will be required to complete two separate forms, it is worth stressing that polytechnic applicants now have to complete a separate form for each polytechnic to which they apply, as well as an UCCA form if they are

also applying to universities. CDP would venture to suggest that two forms will be a marked improvement.

The commitment of CDP to consider at an early stage the possibility of extending PCAS to cover full-time and sandwich degree courses in other local authority institutions is evidence enough that this initiative is not being regarded as the end of the road of rationalizing the applications jungle. But it would be impracticable to attempt to do so all at once. In that context, the initiative was never seen by CDP as a first step to removing the binary line. Others may read into it what they will.

It is a great pity that despite three lengthy telephone calls to the CDP office, and the availability of UCCA and DES press releases, your journal should have so completely misunderstood the nature of polytechnics and of what is being proposed.

M. S. LEWIS

*Committee of Directors
of Polytechnics,
309 Regent Street,
London W1R 7PE, UK*

SIR — The "Binary is divisive" column (*Nature* 2 February, p.400) makes some fair points and the desirability of greater diversity in higher education would be endorsed by many. However, in discussing this matter at least three factors differentiating polytechnics from universities need to be considered and these are, to some extent, run together in the argument presented for scrapping the binary line. The three factors are non-autonomy/autonomy, lower unit costs/higher unit costs and more vocational/less vocational approaches to teaching.

The article does not make clear whether the writer considers that polytechnics should become chartered institutions but there are good arguments that this should happen in some institutions now and in others in the course of time. Discussion of comparative costs (which in practice vary greatly from institution to institution within both groups) tends to be accompanied by a deal of acrimony. Part of the differential relates to salary levels, part to student loads and part to the heavier research involvement of university staff. Unfortunately too much attention is focused on the first of these and indeed, abolition of the binary line tends, in the eyes of many academics in both sectors, mainly to mean unifying of salary structure.

The third factor is in many ways the most important. Many of those who, like myself, have experience of the ways in which polytechnics operate and who value the excellent work done in many departments find it a tragedy that so many polytechnics have moved away from emphasis on vocational and part-time education in the way that the article portrays. What is

needed is a resolve to get back to the founding principles and it does not follow that abolition of the dividing line will produce this. Rather, I suspect, the drift away from doing useful things would be accelerated. There may be another danger here for the universities. Of course much directly vocational work in teaching and research is done, as it always has been, in universities. It would, however, be a tragedy if the vocational test were to be strictly applied to what universities do.

My view is that what the United Kingdom needs is a sober, considered and detached appraisal of what the shape and scope of higher education over the next few decades should be and serious political consideration ought to be given to formulating such a policy as soon as possible.

J. A. BEARDMORE

*Department of Genetics,
University College of Swansea,
Swansea SA2 8PP, UK*

NASA's spending

SIR — I agree wholeheartedly with your argument that the National Aeronautics and Space Administration (NASA) tends to live by pushing for huge new projects and hastily devising rationales for them afterwards (*Nature* 307, 1; 1984). However, your second theme — that the manned station would be an outright disaster for space sciences and astronomy — is a distinctly false charge.

From the start, NASA's design for the station has included a large unmanned platform, carrying astrophysical instrumentation, at a distance of several thousand metres from the manned station. This design, from the point of view of space science, provides the best of both possible worlds. It provides freedom from interference by crew motions or vented-gas contamination (a far more serious shuttle and manned-platform problem which *Nature* failed to mention). But it also allows regular instrument repair and maintenance by a nearby manned crew — a very important factor, as Spacelab-1, Skylab, Solar Max and a legion of other unmanned spacecraft have shown us. In addition, the servicing of other scientific spacecraft has been strongly indicated as a station justification.

Your sneer that the station will serve as a base for research on "whether you can make the perfect ball-bearing in space" also misses the point. From the start, zero-g manufacturing — particularly of drugs — has been recognized as the single greatest justification for a centralized station, manned or unmanned. In sum, the rationale for a manned station remains extremely wobbly — especially as compared to an unmanned station with frequent maintenance-crew visits — but it is far stronger than *Nature* states.

R. BRUCE MOOMAW

*2953 Oakleaf Drive,
Cameron Park, California 95682, USA*

Scientists in Pinochet's Chile

SIR — As scientists working in Chile, we value the concern that *Nature* has shown for the state of academic life under Chile's dictatorial regime in publishing Simon Litvak's commentary (*Nature* 306, 11; 1984). We share Litvak's criticism of the ways in which Chilean universities have been abused for the past ten years. Nevertheless, we think that his article does not fully convey the atmosphere in which Chilean scientists have been living.

Litvak writes that "repression has clearly decreased in recent years" and on this basis he draws an optimistic perspective for the future of Chile's universities. Unfortunately he is wrong. Repression in our country and its universities has lately not decreased but rather dramatically increased, as clearly stated in the 1983 United Nations Report on Human Rights. The Catholic Church, through its Vicaria de la Solidaridad, and the widely respected and politically pluralist Chilean Committee for Human Rights concur with this appraisal. Furthermore, the police continue brutalizing students, as acknowledged by the official press, and paramilitary guards are seen at every gate of the Academia Superior de Ciencias Pedagogicas, severed from the University of Chile in 1981.

	1981	1982	1983
Political arrests	(1) 909 (2) 967	(1) 1213 (2) 1789	(1) 3632 (2) 2824
Internal exile*	(1) 60 (2) 60	(1) 60 (2) 66	(1) 98 (2) 48
Torture denunciations†	(2) 68	(2) 123	(1) 60 (2) 90

(1) Catholic Church, Vicaria de la Solidaridad (9 months)

(2) Chilean Committee for Human Rights (6 months)

* Administrative decision, not judicial.

† Reported personally to one or both (1) and (2), or the judiciary.

As to the "further sign of improvement" that Litvak sees "in the fact that new university legislation committees are now working with impressive freedom", he might have known that the participants were chosen by the government; we can now inform him that the outcome of their labour is a project that further reduces the already scant academic autonomy of Chilean universities. Actually, even the university authorities, also appointed by the government, opposed the project and it has been shelved. The influential association "Andres Bello", to which university professors and dismissed professors from different Chilean universities are affiliated, states that "the withdrawal of this project . . . cannot . . . hide its arbitrary generation nor the conceptual errors it contains". In fact the association had provided the conceptual framework required by legislation concerning the universities but its proposals were ignored. A further example of the lack of dialogue between authorities and university people was the public plea signed by hundreds of professors asking for democratization of universities in order to stop their progress-

ive deterioration. This plea has not been acknowledged by the military government.

Litvak refers "to the better overall functioning" of the Catholic University and implies that this stems from respect due to the Church and also to the permanence of its president, a retired admiral. He fails to mention that Cardinal Raul Silva had to relinquish his role as chancellor to a vice-chancellor who was less outspoken on human rights. He also fails to mention that some influential professors of the Catholic University have contributed to the ideology of the military regime and in particular to its economic policies. For these reasons the Catholic University is indeed a special case and differs from other Chilean universities. Consider, for instance, the case of the University of Chile, which is still the most important. The Association of Professors of its Faculty of Physics and Mathematical Sciences recently said in an open letter to the Minister of Education: "Since the intervention at University of Chile, more than a decade ago, it has suffered unrelenting attacks of a kind unknown in its more than hundred years of life. It [the University of Chile] has been dismembered, turning it from a national university into a local one. For political and/or ideological reasons, many academicians, students and administrative personnel have been persecuted. The university budget has suffered systematic reductions . . . The present situation is such that the academic units that form the University of Chile now receive [funds that are] less than half of what they received in 1973."

We cannot be optimistic while the universities are still under direct military rule. Revival of our universities is difficult to foresee in the general atmosphere in which academic life has taken place in Chile for the past ten years. We hope that these matters will be of concern to scientists and intellectuals everywhere. It is perhaps impossible to quantify the devastating impact on academic life of the take-over of a democracy by a dictatorship but it must be borne in mind by anyone contemplating the future of his own country or the fate of geographically distant societies.

PATRICIO CORDERO, MANUEL A. GARRETON, HUMBERTO GIANNINI, ALEJANDRO GOIC, LUIS IZQUIERDO, RAMON LATORRE, JOSÉ MINGUELL, TITO URETA, FRANCISCO F. VARELA, ENNIO A. VIVALDI
The signatories of this letter are all professors of Universidad de Chile, Casilla 653, Santiago, Chile.

SIR — It is extremely sad that the only thing my colleagues seem to have registered is the moderate optimism I used to analyse the current situation and evolution in Chilean universities. My essential aim was to

convey the sense of intrinsic perversity I have always felt about the military intervention in Chilean universities. I wrote the article during a 2-month stay in Chile last year, after meeting many colleagues and reading a considerable amount of material concerning the situation in that country over the past few years. Having lived in Chile for thirty years, visited the country every year for the past 6 years and lived there recently for a whole year (1980–81), I think I have the right to give a personal view of the situation in that country. As I wrote in my article, the situation in the middle of 1983 was quite favourable for positive evolution since the government was in a defensive position due mainly to the poor economic situation and the successful rallies organized by the opposition forces. Unfortunately, the failure of a crucial meeting last September gave new strength to the Pinochet government. As a result the Minister of Education was changed. While the previous incumbent was in charge of the universities a certain degree of demilitarization of these institutions was observed but it is obvious that the "military lobby" in Chilean universities had gathered enough strength to persuade Pinochet to oust her. It is true that those involved in the formulation of a new university law were named by government officials, but I reaffirm my previous assertion that some of the named participants were known for their dissenting views and that the general atmosphere of those discussions was more democratic than expected. The naming of a new Minister of Education coincided with the shelving of these committees, as stated in the letter of Professor Latorre *et al.* All these unfortunate events, which have slowed the demilitarization process in Chilean universities, happened while my commentary was already in press with *Nature*.

The human rights situation in Chile is far from acceptable, but to deny a certain degree of improvement in recent years is an act of sheer manicheism or bad faith. The possibility of travel abroad without restrictions and the ability to publish dissenting views in several national or international publications are not enjoyed by intellectuals of other dictatorial states.

Finally, I would like to salute the courage of my distinguished colleagues in signing this letter. If repression in Chile is as harsh as they affirm, we should expect a brutal reaction against them from the Chilean authorities. Let us hope that I am right and that they will be able to continue to play a crucial role in forming a generation that will take into their charge, as we deeply hope, the full democratization of the country.

SIMON LITVAK

*Institut Biochimie Cellulaire,
Institut Neurochimie du CNRS,
1 Rue Camille Saint Saens,
33077 Bordeaux Cedex, France*

No patience for the paranormal

Claims that "orthodox" science should pay more attention to paranormal phenomena were, contrary to supposition, heeded seriously a century ago, but to no effect. The onus is now on the unorthodox.

THE University of Edinburgh may not be quite as disreputable as it may seem from having won the competition among British universities for the late Arthur Koestler's legacy for research on paranormal phenomena. In earlier times, Harvard and Stanford Universities and the University of Pennsylvania were all engaged on similar enterprises. So much is clear from a history of psychic research between 1914 and 1939 by Mr Brian Inglis, published this week as *Science and Parascience* (Hodder and Stoughton, £12.95). The remarkable feature of the present time is not that one university should be prepared to spend Mr Koestler's £500,000 and to appoint a professor of parapsychology, but that interest among serious people in the paranormal has been so much less since the Second World War than in its heyday after the First World War.

Why should this be? Part of the explanation may be that suggested by Mr Inglis — the pathetic interest of the relatives of those killed in the huge European battles of 1914–18 in renewing relationships with their dead by means of spirit mediums. But long before that, people such as Sir William Crookes, Lord Rayleigh and even J.J. Thomson in Britain, and William McDougall in the United States, had declared themselves open-minded on the paranormal. Part of Mr Inglis's implied reproach to contemporary science is that so many people have set their faces against such lines of enquiry, usually he says from fear for their academic reputations. What force has that charge?

Although not a scientist, Mr Inglis is well qualified to make this charge. He is a distinguished British (strictly, Irish) journalist who has written widely on matters as different as the Irish problem, and the industrial revolution of the nineteenth century. Partly because of his friendship with Arthur Koestler, he has become an unofficial historian of the paranormal, first with *Natural and Supernatural*, now with the volume published this week.

Inglis explains that he has done his best to write as an historian, not as a "believer", and indeed his book does tell tales of spirit mediums who turned out to be charlatans as well as of those who survived investigation. Mr Inglis's claim to impartiality is weakened by his deadpan description of the work of J.B. Rhine at Duke University in the 1930s; the claim that some people were able to identify marked cards being looked at by a second person

more often than chance would allow has, to say the least of it, been undermined by later demonstrations that the experimental design did not exclude opportunities for collusion, even barefaced cheating. But, mostly, Inglis has no need to bias his own account of the connections between orthodox and unorthodox science; in the inter-war period, many serious people were willing to give time to the paranormal.

Why should that have been the case? Mr Inglis, starting from the position that it is wayward to the point of irresponsibility that contemporary science is habitually scornful of claims of paranormalcy, is understandably uninformative on this point. But that late-Victorian science should have thrown up people prepared to devote themselves to the investigation of these phenomena is not surprising. With the systematic study of cognitive processes (psychology) in its infancy, and with even orthodox science littered by unexpected phenomena, it is understandable that people should have paid some attention to the popular craze for attending seances at which mediums would make voices materialize from the dead and, often, quantities of a material called ectoplasm that would from time to time assume the form of a human hand or head.

In 1876, William Barrett, professor of physics at the Royal College of Science in Dublin, even went so far as to hold forth on the subject at the annual meeting of the British Association, that year at Glasgow. This journal at the time reported that "an excited discussion arose. Many phenomena of mesmerism, clairvoyance and spiritualism were alleged, and Mr Crookes (later Sir William), Mr (A.R.) Wallace, Lord Rayleigh and Dr Carpenter (W.B. Carpenter, the palaeontologist) expressed opinions which are well known, based on facts witnessed by themselves." Barrett's enthusiasm, apparently undiminished by the hostility of the sceptics, led a few years later to the formation of the Society for Psychical Research to put psychic phenomena through more rigorous tests than had previously been carried out. The president of the new society was Henry Sidgwick, professor of philosophy at the University of Cambridge, England, whose opposite number at Harvard University, William James, soon became president of the comparable society in the United States.

To begin with, the paranormalcy people were thus flush with eminent supporters, so

that it was also natural that gifts of money should be thrust at often-reluctant universities for the support of research in the field. Stanford accepted \$50,000 for a fellowship in telepathy only because the donor was the brother of the founder, and then took care to appoint a sceptic to the post. Inglis wrings his hands over this development, as well as over the growth of scepticism in the Society for Psychical Research, quoting in his support the remark of A.R. Wallace that the society existed "for the suppression of facts, for the wholesale imputation of imposture, for the discouragement of the sensitive and the repudiation of every revelation of the kind which was said to be pressing itself upon humanity from the regions of light and knowledge".

But what else can Mr Inglis and his fellow-thinkers expect? By the evidence of his own book, there was a time almost within living memory when academic people paid close attention to claims that paranormal phenomena occur, and were discovered to have wasted their time unmasking fraud. In 1926, this journal, seeking to end a long and acrimonious correspondence on the subject, acknowledged that there was a need "that a competent body should exist for making out a *prima facie* case for any new phenomena" (*Nature* 118, 721; 1926) and went on to suggest that the Society for Psychical Research might with advantage be combined in that role with the then new "National Laboratory for Psychical Research" (proprietor one Harry Price). Mercifully, the advice was not followed, for Mr Price was himself eventually unmasked as the principal organizer as well as investigator of the notorious Borley Rectory poltergeists. And the record of charlatanry has continued ever since.

This is why it is mystifying that people like Mr Inglis should still insist that "orthodox" science blindly refuses to enter a whole field of science which is supposedly ripe for investigation but which is never defined with any clarity. Yet, in this book as in its companion literature, the argument that serious people should spend more time looking into probably spurious happenings is bolstered by the argument that quantum mechanics "has continued to provide evidence of paranormal-type phenomena, translocation and action at a distance at the micro level". If the question is so simple, why are verifiable phenomena so few?

John Maddox

Particle physics

Can the gallium detector solve the solar neutrino problem?

from W. Hampel

SOLAR neutrino experiments, designed to detect the neutrinos generated in the nuclear fusion processes in the Sun, are among those difficult experiments that aim to detect very rare events. The only type of detector able to overcome the huge background problems in detecting these neutrinos is the radiochemical detector. An important quantity for radiochemical experiments is the neutrino capture cross-section for the target nucleus. A recent paper of Orihara *et al.*¹ is concerned with the cross-section of the gallium detector, upon which rest current hopes of resolving the discrepancy known as the solar neutrino problem.

The only completed experimental attempt to detect solar neutrinos, the Brookhaven Chlorine Experiment², has resulted in a neutrino signal of 1.8 ± 0.3 SNU (1 SNU = 1 neutrino capture per second in 10^{36} target atoms). This is almost four times lower than the 6.9 SNU predicted by the so-called Standard Solar Model (SSM)^{3,4}. It is this discrepancy that has long been known as the solar neutrino problem.

Because of the energy threshold (814 keV) of the $^{37}\text{Cl}(\nu, e^-)^{37}\text{Ar}$ reaction on which the chlorine detector is based, it is mainly sensitive to the ^7Be (862 keV) and ^8B (0–14 MeV) neutrinos. While contributing 75 per cent of the expected rate in this experiment, ^8B neutrinos are produced in a very rare side branch of the solar fusion chain and are insignificant in the total energy production in the Sun. Unfortunately, the chlorine detector cannot respond to the bulk of solar neutrinos, the pp neutrinos (0–420 keV), generated in the primary fusion reaction.

It has long been recognized that an experiment with a threshold low enough to detect the pp neutrinos might, among other things, provide the answer to the problem imposed by the chlorine experiment. The only experiment of this type which has been demonstrated to be feasible is the gallium experiment⁵, based on the neutrino capture reaction $^{71}\text{Ga}(\nu, e^-)^{71}\text{Ge}$ (236 keV threshold). The SSM prediction for that experiment³ is 106 SNU (pp: 66 per cent, ^7Be : 27 per cent). Even if one assumes that the low chlorine result is due to some unknown problems with the solar model, the most probable result for the gallium detector will still be near 85 SNU.

The only other seriously discussed explanation for the solar neutrino problem is a reduced electron neutrino flux at the detector due to neutrino oscillations; mixing of the electron neutrino with neutrinos of other flavours (oscillations) may occur if neutrinos are not mass eigen-

states. In this case the expected neutrino signal for the gallium detector would be distinctly reduced and certainly less than 70 SNU. That is why the gallium detector should be able to decide whether the explanation of the solar neutrino problem is that there are serious problems with the solar model or that neutrinos oscillate. The latter would involve a non-zero rest mass of the neutrino, which would have severe consequences for elementary particle physics and, possibly, for cosmology.

The interpretation of the result from a gallium detector requires knowledge of the cross-section for neutrino capture in ^{71}Ga . This cross-section is dominated by transitions to the ^{71}Ge ground state, which can be reliably calculated from the observed β -decay between the two states⁶. The rates given above for the gallium detector are based on this ground state cross-section. However, there may be additional contributions from two states in ^{71}Ge at excitation energies 175 and 500 keV which may be populated in particular by ^7Be neutrinos. In order to calculate the cross-section leading to excited states one needs the β -decay ft-values (Gamow-Teller strength) of the respective states. In principle this information can be obtained from the systematics of β -decay or from (p,n) forward scattering experiments since the (p,n) charge-exchange reaction connects the same states in the target and product nuclei as neutrino capture or (in inverse direction) β -decay.

Orihara and co-workers¹ have measured the $^{71}\text{Ga}(p, n)$ reaction at 35 MeV proton energy and have extracted from their data ft-values corresponding to transitions from the ^{71}Ga ground state to the ^{71}Ge states at 175 and 500 keV. Their results yield a total rate of 135 SNU for the gallium detector, 29 SNU larger than the predicted capture rate for the transition to the ground state alone (the number of 149.6 SNU quoted in their

paper is incorrect because of a computational error). The increase is dominated by an increased ^7Be fraction in this rate (pp: 53 per cent, ^7Be : 36 per cent). This causes the prediction for the gallium detector to become more dependent on details of the solar model. It will not, however, qualitatively affect the potential of the gallium experiment to provide the answer to the solar neutrino problem.

In addition, it should be noted that there is a problem connected with the interpretation of the (p,n) data at 35 MeV in terms of β -decay ft-values: there could be contributions to the (p,n) cross-section which do not simply correspond to allowed Gamow-Teller transitions. Consequently, the neutrino capture cross-section extracted from (p,n) measurements would be overestimated. Indications that this may be the case come from a preliminary result of 14 SNU for the excited state contribution in a (p,n) experiment on ^{71}Ga using 120 MeV protons at the Indiana University cyclotron (J.N. Bahcall, personal communication).

Since there will always remain some ambiguities in the interpretation of (p,n) data in terms of neutrino capture cross-section, it is important to note that an experiment with an artificial ^{51}Cr neutrino source planned as an integral part of the gallium experiment⁵ will eventually be able to answer the question about excited state contributions. ^{51}Cr emits neutrinos of 746 keV, sufficient in energy to populate the 175 and 500 keV levels in ^{71}Ge . The ^{51}Cr experiment will thus provide not only a test of the entire detector system but will also yield direct information on the ^{71}Ga neutrino capture cross-section relevant for solar neutrinos. □

W. Hampel is at the Max Planck Institute for Nuclear Physics, D-6900 Heidelberg 1, Postfach 10 39 80, Saupfercheckweg 1, FRG.

1. Orihara, H. *et al.* *Phys. Rev. Lett.* **51**, 1328 (1983).
2. Davis, R., Cleveland, B.T. & Rowley, J.K. *Science Underground, AIP Conf. Proc.* **96**, 2 (1983).
3. Bahcall, J.N., Huebner, W.F., Lubow, S.H., Parker, P.D. & Ulrich, R.K. *Rev. Mod. Phys.* **54**, 767 (1982).
4. Fowler, W.A. *Science Underground, AIP Conf. Proc.* **96**, 80 (1983).
5. Hampel, W. *Science Underground, AIP Conf. Proc.* **96**, 88 (1983).
6. Bahcall, J.N. *Rev. Mod. Phys.* **50**, 881 (1978).

100 years ago

THE REMARKABLE SUNSETS

THOUGH we are no longer favoured with the gorgeous sunsets which marked the autumn and early winter, yet two phenomena are still frequently visible which seem referable to the same cause as those splendid displays.

The first is the unusual *white glow* in the western sky before sunset which was an almost constant precursor of the brilliant and long-continued colouring of the past months. It was very marked on November 8, the occasion of the first remarkable sunset, and it still to be seen on almost any fine evening before the sun sets, though it is no longer followed by the more striking phenomena.

The second is a decidedly *unusual pink tinge*

occasionally visible for some ten to twenty degrees round the sun when shining in a somewhat hazy sky, the colour being brought out with great distinctness if light cumulus cloud happens to be passing across it. I first observed it about 1 p.m. on Sunday, March 2, and it was very marked last Thursday (20th) between 10 and 11 a.m., and again on Friday (21st) between 1 and 2 p.m., as well as on one or two other days which I have not specially noted.

May not both be due to the gradual subsidence to a lower level in our atmosphere of the particles which at a higher elevation caused the wonderful colouring of the past months?

Hampstead, March 24 B. W. S.
From *Nature* **29**, 503, 27 March 1884.

Oxford physics

Opportunity lost in 1865?

from N. Kurti

ADAGES like "Oxford is for manners and marmalade, Cambridge for sausages and sums" take a long time to die. The sight of Frank Cooper's marmalade factory near the station has long ceased to gladden the heart of the visitor to Oxford, yet the myth still persists that Oxford is fine for the humanities but for science, engineering or mathematics you had better go to Cambridge. Several reasons can be given for this view, such as the renowned and large engineering school in Cambridge — in 1945 the ratio of Cambridge to Oxford engineering intakes was 10 to 1, and even today it is somewhat above 2 to 1 — and the multidisciplinary nature of the Cambridge Natural Sciences Tripos which was, and perhaps still is, the reason for the high proportion of Cambridge graduates among science teachers. But there is another explanation which so far does not seem to have been considered.

During the 50 years 1870–1920, Cambridge justly earned the reputation of being the centre of excellence for physics by virtue of the discoveries made in the Cavendish Laboratory or by Cavendish men. During the same period, the Clarendon Laboratory in Oxford remained practically unknown, and this almost total disregard of physics research coloured views about the standing of science in the oldest British university.

Why was Oxford so backward? The answer is obvious: during the period in question, the Cavendish Professors were Maxwell, Rayleigh and J.J. Thomson, three household names, while very few, even among physicists, have heard of R.B. Clifton who was Dr Lee's Professor of Experimental Philosophy and head of the Clarendon Laboratory from 1865 until his retirement in 1918 at the age of 81.

Curiously, Clifton had his application for the Oxford chair privately printed, which is fortunate since there is nothing in the university archives or in the minutes of the Hebdomadal Council about his election. In support of his application, Clifton cites two published papers and prints a remarkable collection of 20 testimonials from eminent people, including J. C. Adams, Bunsen, Joule, Kirchhoff, Stokes and W. Thomson (later Lord Kelvin). All referees emphasize the excellence of Clifton's lectures, but their comments on his ability to conduct original research — based on one single paper published jointly with Professor Roscoe, his colleague at Owens College, Manchester (now the University of Manchester) — are vague and lukewarm. This is all the more significant since testimonials addressed to and seen by the candidate are likely to be laudatory. Never-

theless the electors chose Clifton and the consequences of their decision may be gauged by the following quotation from the Royal Society's obituary notice, which appeared soon after Clifton's death in 1921:

The design of the [Clarendon] Laboratory was his first task. . . . Much of this was designed and redesigned by him until perfection, or something approaching it, was reached, and so much loving care had been spent on an instrument that it needed to be kept jealously under lock and key, taken out from time to time to be dusted and cleaned, possibly to be used in lecture, but entrusted never to the careless handling of a student of Physics.

Of research work there was little: the Laboratory was intended for teaching. . . . Almost the only great piece of research appearing from the Clarendon Laboratory in his time was Boys' determination of the constant of gravitation [carried out in the vibration-free cellar of the Clarendon rather than in the Royal College of Science in South Kensington where C.V. Boys worked].

Clearly, Clifton did not fulfil the expectations placed in him and the choice was unfortunate, even bizarre, in view of the widespread belief in Oxford that Hermann v. Helmholtz of Heidelberg, who was already a scientist of world fame, had been an applicant for the post.

I have heard one ingenious and plausible explanation why Oxford preferred Clifton to Helmholtz. There was, in the 1860s, a passionate debate in the university about whether Oxford should continue to be primarily concerned with teaching the young or whether it should imitate the great continental universities and become a

centre for learning and research as well. One of the advocates of the latter view was a distinguished scholar of German origin, Max Müller, Professor of Comparative Philology, and, so the story goes, the Oxford 'establishment' felt that it was bad enough to have one famous German professor trying to change the university — it would be unwise to have a second.

But did Helmholtz really apply? The university archives say nothing about the election. Koenigsberger's biography of Helmholtz says that, in March 1864, Helmholtz visited Oxford for two days, and a letter to his wife contains thumbnail sketches of Max Müller, with whom he was staying, and of the university:

Oxford is perhaps unique of its kind; its many old, beautiful and well-preserved buildings are splendid and are evidence of immense wealth. . . . I now understand the devotion of an Englishman to his University. The system works admirably for the education of 'gentlemen' but it cannot lead to much in science and learning and it needs a very strong interest in learning to prevent a Fellow from sinking into idleness. Max Müller is perhaps the only one doing any work here.

But it is a letter written from Paris on 11 April 1866, five months after Clifton's election, which gives the clue to the story about Helmholtz's candidature:

At eleven o'clock I had to be back to lunch with M. Hermite and the mathematician Prof. Smith from Oxford. It was mentioned that they had wanted to appoint me as Professor of Physics in Oxford but they could raise only £700 salary which, admittedly, is more than we have in Heidelberg, but hardly enough to live in the same comfort in England. . . . I think, therefore, that it was quite correct for Professor Max Müller to assure them that I would certainly not come on those terms. . . .

Why did Max Müller prevent the invitation to Helmholtz? He knew Helmholtz well enough to ask him informally whether



Would H. v. Helmholtz (left) have succeeded where R.B. Clifton (right) failed?

he would at all consider an official invitation. The salary was not insultingly low; when in 1871 Sir William Thomson invited Helmholtz to become the first Cavendish Professor in Cambridge the stipend he could offer was not much more than £700. A possible — and to my mind the most probable — explanation for Max Müller's declining the invitation on behalf of Helmholtz without consulting him is that, at the time, Max Müller was regarded as perhaps the most distinguished and internationally best known scholar in Oxford, and he may have felt that Helmholtz would outshine him. He effectively vetoed the invitation to avoid this risk.

Would Helmholtz have accepted the chair? It appears at first sight that he approved of Max Müller's action but the relevant passage in the German original (*"Ich finde es also ganz in der Ordnung . . ."*) sounds less definite than the English version. Then there is the passage omitted by Koenigsberger in the quotation above and marked differences between his version and the rendering of the same letter in the biography of Anna von Helmholtz (Helmholtz's wife) by her daughter, Ellen von Siemens-Helmholtz. The passage in the Koenigsberger version "but they could raise . . . not come on those terms" appears in the Anna von Helmholtz biography as "and that Professor Max Müller had told them assuredly that I would not accept the offer". The original of the letter cannot be found. The Archives of the Akademie der Wissenschaften in East Berlin has only Helmholtz's scientific papers and correspondence. The personal papers remained originally in the family but were transferred to a green safe at the academy which survived the allied bombing attack of 3 February 1945 and was still in its place in 1947. Subsequently the safe was opened, its contents disappeared and we shall probably never know the exact wording of Helmholtz's letters.

One can only speculate whether, but for Max Müller, Helmholtz would have become Professor of Experimental Philosophy in Oxford, how the Clarendon Laboratory would have fared under him and whether 50 years of neglect would have been avoided. For the sad truth is that Clifton's lofty aim "to aid . . . as far as in me lies, in securing and maintaining for the University a reputation in science befitting the dignity of Oxford" remained unfulfilled.

In preparing this article I was greatly helped by the late Professor W. Haberditzl (Humboldt Universität), by Dr Christa Kirsten (Director of the Archives of the Akademie der DDR) and by Dr and Mrs Hermann von Siemens (München). □

N. Kurti is Emeritus Professor of Physics now in the Department of Engineering Science, Parks Road, Oxford OX2 3PJ.

Low-temperature physics

Persistent currents in liquid ^3He

from Peter McClintock

If you once start liquid ^3He moving round a circle, it will continue to flow indefinitely without any further assistance and without slowing down at all, according to a recent experiment at Cornell University. The work in question, which was carried out by P.L. Gammel and J.D. Reppy of Cornell in collaboration with H.E. Hall of the University of Manchester, is reported in *Physical Review Letters* (52, 121; 1984).

A feature that liquid ^4He , liquid ^3He and the electron gas in certain metals have in common is that, as a result of a low-temperature transition, they seem to lose all their viscosity and are then able to flow without friction and thus without any dissipation of energy. The critical temperatures

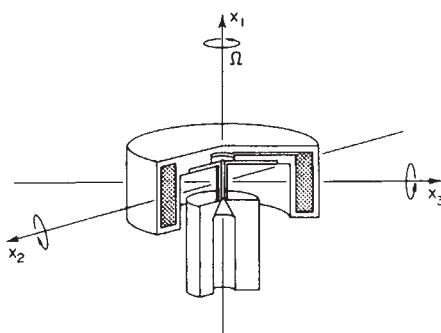


Fig.1 Schematic diagram of the ac gyroscope. The ^3He fills the shaded cross-section of the cap of the mushroom via the hollow stem. Drive and detection electrodes (not shown) oscillate the mushroom about any of the axes X_1 , X_2 and X_3 , and monitor the amplitude of oscillation. By movement of the entire cryostat the mushroom can be rotated at a steady angular velocity Ω about axis X_1 .

for the onset of this so-called superfluid behaviour ranges from only 2 mK in liquid ^3He to about 20 K in a metallic alloy. In each case, below the transition temperature the fluid seems to divide into two distinct but freely intermingling fractions: a normal fluid component, much like any other liquid, and a superfluid component that exhibits the dissipationless flow properties.

One of the best ways to demonstrate that a superfluid flows without friction is to show that, once it has been set to flow around a closed circuit, it keeps going without any diminution of velocity. (The flow of a normal liquid in such an experiment would gradually slow down.) However, whereas it is relatively easy to demonstrate persistent currents in a superconductor, it is not so simple in superfluids.

For a superconductor, if one joins the two ends of a coil of superconducting wire and starts an electrical current flowing

round the coil, both the presence and the strength of the current can readily be measured in terms of the magnetic field that is created. An experiment of this type has been run for more than a year without any measurable change occurring in the field produced by the circulating current.

A persistent flow current of superfluid ^4He is far more difficult to detect because the atoms, being neutral, create no associated magnetic field by their movement. Nevertheless, about fifteen years ago, J.D. Reppy of Cornell University developed a cunning technique capable of detecting the presence of persistent flows of this nature, based on what he called a superfluid gyroscope. His method exploits the propensity of gyroscopes to try to move at right angles to the direction in which they are being pushed. Thus, in a 'gyroscope' of superfluid flowing round a sensitively suspended torus, the existence of superflow is readily demonstrated if, upon turning the torus slightly about one diameter, it responds by trying to turn about another diameter, at right angles to the first. The Reppy gyroscope provides an outstandingly successful method for studying persistent currents in superfluid ^4He . It is much less easily applied to superfluid ^3He , partly because of the relatively small amount of liquid that can be cooled to the exceedingly low temperatures that are required.

Most people would have guessed that a superfluid ^3He gyroscope would, in practical terms, be quite impossible to set up. Nonetheless, what Gammel *et al.* have now devised — and operated successfully — is nothing less than a modified form of the original Reppy gyroscope. The strange instrument that they have created (see Fig.1), which they refer to as an 'ac gyroscope', consists, in essence, of a hollow 'mushroom' in which the head is made of epoxy and the (narrow) stem is made of springy beryllium-copper. The mushroom can be oscillated at small amplitude about any of the three axes X_1 , X_2 or X_3 , and the amplitudes are detected by sensors. The cap of the mushroom is filled with liquid ^3He through its hollow stem and the whole cryostat, which refrigerates the mushroom, is rotated (relatively rapidly) for a while about axis X_1 , to try to set up a persistent current of ^3He around the circumference of the cap, and then brought to rest. To test whether a current persists, the mushroom is then oscillated at its resonant frequency about axis X_2 ; if there is a persistent current of ^3He , the gyroscope effect will cause, in addition, a small oscillatory motion about the axis at right angles, X_3 . That is precisely what Gammel *et al.* have observed for the

B-phase of superfluid ^3He . They find that there is no observable decay of the persistent current over a period of an hour and have been able to deduce, on this basis, that the viscosity (if any) of the superfluid component is less than that of the normal fluid component by at least 10^8 — an enormous factor.

This is a very satisfying result, particularly in view of some aspersions previously cast on the superfluidity of the B-phase. Eisenstein and Packard (*Phys. Rev. Lett.* **49**, 564; 1982) had reported the observation of some dissipation apparently associated with the flow of ^3He .

B through a narrow channel connecting two reservoirs. Although Brand and Cross subsequently proposed a plausible explanation of these effects in terms of events occurring in the reservoirs, rather than in the channel (*Phys. Rev. Lett.* **49**, 1959; 1982), there remained a lurking suspicion that perhaps ^3He -B was, in some way, not quite such a 'good' superfluid as ^4He or the electron gas in a superconductor. These residual doubts can now be laid to rest. □

Peter McClintock is in the Physics Department, University of Lancaster, Lancaster LA1 4YB.

X-ray astronomy

Milestone for EXOSAT

from Mike Watson

EXOSAT, the European Space Agency's first satellite for X-ray astronomy, was launched in May last year and passed a second milestone in mid-December when the early scientific results were publicly presented at a special one-day session of a meeting at Florence*. The early phase of the EXOSAT mission was not free from problems, the sad demise of two of the four imaging X-ray detectors, for example, but also provided some pleasant surprises; the background levels for both the medium-energy experiment and the gas scintillator proportional counter were substantially below those expected. Judging by both the quality and quantity of results presented at the meeting in more than thirty papers, EXOSAT should fulfil its role of providing a major X-ray observational facility for Europe during the next three years.

EXOSAT has several novel features which give it important advantages over previous X-ray astronomy missions, as Peacock (European Space Research and Technology Centre, ESTEC) described. In particular, the sensitivity of the channel multiplier array imaging detectors in the soft X-ray/extreme UV permits studies of X-ray emissions in the 10–300 Å range — a part of the spectrum hitherto barely explored. Another important feature is the satellite's highly eccentric 90-hour orbit, which makes possible long uninterrupted observations important for the study of variability on time scales from hours to days characteristic of many galactic X-ray binaries. Real-time operation of the spacecraft from the European Space Operations Centre (ESOC) control centre in Darmstadt allows the entire payload to be used as a true X-ray observatory. This has proved enormously valuable in programmes coordinated with ground-based or International Ultraviolet Explorer studies and also allows EXOSAT to respond quickly to new phenomena such as X-ray transient outbursts. (60 per cent of

observations made with EXOSAT in the first five months took advantage of this capability.)

At the Florence meeting, the emphasis was on studies of galactic objects, particularly binary X-ray emitters of all types. The most dramatically variable objects in the X-ray sky are the X-ray transients which undergo intense outbursts paralleling those of classical novae.

Blisset (ESOC) and Peacock described two X-ray transients (4U1543–47 and V0332+53) which conveniently flared up during the first six months of the EXOSAT mission. X-ray outbursts from both sources were first noted by the Japanese Tenma satellite, but the flexibility of EXOSAT made it possible to obtain within a few days X-ray positions accurate enough for the optical counterparts of both transients to be quickly identified. Observations of the transient V0332+53 also revealed a coherent 4.4-s pulsation together with chaotic flaring activity down to millisecond time scales, behaviour characteristic of the black hole candidates Cyg X-1 and GX339–4. This observation may be especially important for our understanding of the black hole candidates, for the coherent 4.4-s pulsations are almost certainly the signature of a rotating magnetized neutron star.

Among other results on bright galactic X-ray sources, Parmar (ESOC) and Kahabka (Max Planck Institute, Garching) reported that the binary Her X-1 had undergone a transition into a low-flux state in which its normal X-ray characteristics seem to be absent, even though X-ray heating of the optical companion seems still to be taking place. Pietsch (Max Planck Institute, Garching) had a similar story to tell about 4U2129+47, which was undetectable during the EXOSAT observation.

The continuous 30-h EXOSAT observation of Cyg X-3 described by Willingale (University of Leicester) and Blisset was particularly impressive. The wealth of new

features discovered in the X-ray light curve will be a challenge for models of this exotic system. X-ray bursters and bulge sources also featured, with presentations by Sztajno (Max Planck Institute, Garching) of the discovery of absorption dips in 4U1755–33 and by Turner (University of Leicester) of an unusual absorption feature in X-ray spectra of several bursts from 4U1636–536 obtained with the medium-energy experiment which could be due to iron absorption red-shifted by the strong gravitational field near the neutron star.

EXOSAT is also providing valuable data on cataclysmic variables, interacting binary systems involving accretion onto a white dwarf companion. Current interest centres on those systems, known as 'polars' and 'intermediate polars', containing a magnetized white dwarf with a field strong enough to ensure that accretion takes place primarily onto the magnetic polecaps.

Heise (Space Research Laboratory, Utrecht) announced the discovery of completely anomalous behaviour in the best-known polar, AM Her; the modulation of the soft X-ray flux at the 3.1-h binary period is almost exactly out of phase with the hard X rays, perhaps implying emission from two active magnetic poles. For EX Hya, a possible intermediate polar, Cordova (Mullard Space Science Laboratory/Los Alamos) showed the clear presence of a 67-min period together with a weaker modulation at the 98-min binary period in the soft X-ray light curve. Pietsch described observations of the established intermediate polar H2252–035 which suggest a 30 per cent orbital modulation of the X-ray flux as well as the well-known 805-s pulse period. Watson (University of Leicester) announced the discovery, during an optical outburst, of X-ray pulsations from the old nova GK Per, with a period of 351 s, which implies that the accreting white dwarf in this system must also have a substantial magnetic field and be yet another intermediate polar. The relative amplitude of the X-ray pulses from GK Per does not change as the mean X-ray flux varies, which may be an important clue to the geometry of the region at the magnetic polecap emitting X rays, as it is difficult to model this behaviour unless the scale-height of this region above the white dwarf surface is small.

EXOSAT's low-energy imaging detectors can be used in conjunction with filters to provide X-ray 'colours' in the important spectral region spanning the soft X-ray and extreme ultraviolet; this capability is particularly valuable for observations of single stellar objects, as was demonstrated by McKechnie (University of Leiden) and Heise. Heise also presented objective grating spectra of the hot white dwarf HZ 43 and AM Her which indicated the great potential of this instrumental combination. Some of the visually most impressive results were the images obtained of supernova remnants, and the value of spectro-

* The Seventh European Regional Astronomy Meeting was held in Florence on 12–16 December 1983.

scopic observations of these objects was shown by Smith (ESTEC) who presented the first medium-energy spectrum of the supernova remnant W49B. The spectrum is dominated by a remarkably strong iron line with equivalent width 2 keV which may imply non-equilibrium ionization conditions within the X-ray plasma.

By contrast with the attention given to galactic X-ray sources, extragalactic astronomy formed a smaller part of the programme, partly because the problems encountered with the imaging detectors have meant extensive rescheduling of observations but also because of the need to build up a good understanding of the background level in the medium-energy experiment before analysing data from weak extragalactic sources. Nevertheless, interesting imaging observations of 3C120 (Maraschi, University of Milan), BL Lac

objects (Maccagni, University of Milan) and flat-spectrum radio sources (Biermann, Max Planck Institute, Bonn) were presented, while Pounds (University of Leicester) showed the first X-ray spectra of active galaxies obtained with the medium-energy experiment and Branduardi-Raymont (Mullard Space Science Laboratory), the first EXOSAT images of clusters of galaxies.

The early part of the EXOSAT mission has clearly been a success. The range of topics covered at Florence underlines the relevance of X-ray astronomy to many areas of astrophysics, while the enthusiastic participation at this meeting surely points to the rapid growth of this field in Europe. □

Mike Watson is in the X-ray Astronomy Group at the University of Leicester, University Road, Leicester LE1 7RH.

Palaeohydrology

Clues to past climate in river sediment

from Peter D. Moore

It is estimated that 40,000 km³ of water, carrying perhaps 15,000–20,000 million tonnes of suspended material, run off the Earth's continents each year¹. Global figures fail to indicate, however, the very high level of variation found under a range of local conditions. Sediment yield varies with the ease with which underlying geological materials are eroded, with topographical relief, which controls the energy of moving waters, with climate, which influences the total amount and temporal distribution of water movements and with the vegetation, which varies in its capacity to stabilize a substrate. So palaeoenvironmentalists have often looked to river alluvium as a source of information on changing conditions in the past^{2,3}.

Sedimentation rates and the textural quality of sediments vary with conditions in the catchment, but it is difficult to identify the cause of change because of the complex interactions of many factors within the environment. Probably the most influential factor subject to short-term variation is the structure of vegetation, which is obviously very sensitive to human land use. Sundborg¹, for example, quotes data from Tanzania where, in areas of ungrazed scrub vegetation on a 3.5° gradient, only 0.4 per cent of the total rainfall is lost by runoff and erosion is effectively zero. If this vegetation is replaced by grass, there is still no erosion but runoff rises to 1.9 per cent of incident rain. With arable agriculture, in the form of maize, however, 26 per cent of the rain is lost as surface runoff, carrying with it 78 tonnes of sediment per hectare of catchment; and in fallow arable, 50 per cent of the rain is lost and 146 tonnes of

sediment eroded per hectare. Such figures are informative if land use, and therefore vegetation, is the only variable under consideration.

Variation in climate will have a further pronounced effect on runoff and sediment yield, as has been shown by the recent work of Finley and Gustavson in Texas⁶. They have been particularly interested in the influence of storms. In a study area where mean annual precipitation is about 400–500 mm yr⁻¹, a single thunderstorm in May 1978 brought 71 mm of rain in three hours, with a maximum intensity of 64 mm h⁻¹ over a half-hour period. The resulting erosion led to a downcutting of the sandy soils at one point by 62 mm and the material was redeposited further down the valleys in the form of sand and gravel bars and fans. From exposed sections of similar alluvium in canyon walls it was possible to estimate an overall deposition rate of 3 m in 2,100–2,600 yr, but clearly deposition had not been even. A gravel lens of 10 cm thickness found within the sequence could easily have been deposited during a single storm.

Observations of this type serve as cautionary tales for geoarchaeologists attempting to interpret a sequence of alluvial sediments. A recent example is the work of Butzer, Miralles and Manteu⁷ at the site of Alzira, near Valencia in Spain. In the eleventh-century, the site was an Islamic city which was almost completely surrounded by a meander of the Rio Júcar river, which has now been cut off and remains as an infilled oxbow lake. The basal layers of the excavation, at a depth of 5 m, are clayey silts, which contain very little sand and are archaeologically sterile. Above 450 cm, the sand content increases grad-

ually to about 25 per cent at 270 cm, above which there is a further abrupt increase, together with evidence of dwelling construction. Yet more sand builds up above about 150 cm depth.

To assist in the interpretation of the deposits, modern samples were collected from a variety of local sedimentary environments with known depositional characteristics and were compared with the fossil samples in terms of particle size and sorting parameters. In this way, the sedimentary history of the site has been reconstructed.

At the base, conditions were of a low-energy type, with steady deposition and no evidence of erratic floods. Regular flooding occurred during the mid-eleventh century, coincident with the building of the city walls, culminating in high-energy floods in the late eleventh century. Sudden changes in the sand content above about 270 cm are due to further construction at the site. The dwelling had to be abandoned eventually, however, because of catastrophic floods in the sixteenth and seventeenth centuries, leading to the deposition of the sand-rich sediments at a depth of 100–150 cm.

The story, therefore, is one of stable low-energy conditions being replaced in the eleventh century by increasingly severe floods which probably continued into the twelfth and thirteenth centuries. Subsequently, there is evidence for infrequent, but catastrophic, flooding. The question remains, however, whether the changing hydrological circumstances resulted from alterations in human land use or from anomalies of climate.

The fact that, after the eleventh-century floods, conditions never fully reverted to the stability of the tenth century, has led Butzer and his co-workers to suggest that a permanent change in the nature of the catchment had been brought about, most probably because of deforestation during the Islamic occupation of the area. The more recent incidence of irregular catastrophic floods is accounted for in terms of climatic anomalies whose effects have been amplified by the unstable state of the floodplain, which was the result of both the earlier deforestation, itself, and consequent aggradation in the upper part of the valley. □

Peter D. Moore is Senior Lecturer in the Department of Plant Sciences, University of London King's College, 68 Half Moon Lane, London SE24 9JF.

1. Sundborg, A. *Nature and Resources* (UNESCO) 19(2), 10 (1983).
2. Vita-Finzi, C. *The Mediterranean Valleys* (Cambridge University Press, 1969).
3. Dennell, R. W. in *Desertification and Development* (eds Spooner, B. & Mani, H.S.) 171 (Academic, London, 1982).
4. Moore, P. D. *Quat. Stud. Poland* 4, 97 (1983).
5. Gregory, K. J. *Background to Paleohydrology* (Wiley, Chichester, 1983).
6. Finley, R. J. & Gustavson, T. C. *Earth Surf. Processes and Landforms* 8, 63 (1983).
7. Butzer, K. W., Miralles, I. & Mateu, J. F. *J. archaeol. Sci.* 10, 333 (1983).

Oncogenic intelligence

From *c-fos* to *v-fos*

from Inder M. Verma

THE discovery of kinship between retroviral oncogenes (*v-onc*) and their cellular homologues or proto-oncogenes (*c-onc*) has caused much of the recent excitement in cancer research. The precise mode of acquisition of cellular sequences by retroviruses remains obscure, but clearly it is an exceedingly rare event. To date nearly 20 unique *v-onc* genes have been isolated and characterized extensively. One such oncogene, *fos*, has been identified as the transforming gene of FBJ murine osteosarcoma virus (FBJ-MuSV) which was isolated from a spontaneous tumour in a CFI mouse^{1,2}.

FBJ-MuSV proviral DNA contains 4,026 nucleotides which include two long terminal repeats (LTRs) of 617 nucleotides each, 1,639 nucleotides of acquired cellular sequence (*v-fos*) and a portion of the *env* gene encoding the viral envelope protein³ (see Fig. 1). The acquired sequences encode a *v-fos* protein of 381 amino acids, with a calculated molecular weight of 41,601. In cells transformed by FBJ-MuSV, a phosphoprotein of apparent molecular weight 55,000 (on SDS-PAGE) has been identified as the transforming protein⁴. The discrepancy in size is probably due to both phosphorylation and the unusual amino acid composition of the *fos* protein (10 per cent proline) which enables it to be extensively folded.

Using *v-fos*-specific sequences, *c-fos* genes have been isolated from mouse and human cells and their complete nucleotide sequences determined^{3,5}. The predicted mouse and human *c-fos* proteins differ in only 24 of their 380 amino acids. The sequences in *c-fos* that are homologous to *v-fos* are interrupted by four regions of non-homology, three of which represent *bona fide* introns. The fourth, 104-nucleotide, region, which is present in both mouse and human *c-fos* genes, is not an intron but part of the *c-fos* coding sequence; nevertheless it has evidently been deleted during the genesis of the *v-fos* gene with a consequent shift in the reading frame. The result is that all but 5 of the first 332 amino acids of the *c-fos* and *v-fos* proteins (380 and 381 amino acids, respectively, in total) are identical but, because of the out-of-frame deletion, their C-termini are quite different (Fig. 1a).

FBJ-MuSV presumably arose by recombination between *c-fos* sequences and FBJ murine leukaemia virus (FBJ-MuLV), probably through a 5-nucleotide sequence they share at a position corresponding to the 5' end of *v-fos* and 10 out of 11 nucleotides they share at the 3' end (Fig. 1b). Sequences involved in recombination at the 5' end lie in the untranslated region of FBJ-MuLV and

immediately downstream from the predicted 5' end of *c-fos* mRNA. At the 3' end, they involve the p15E region of the FBJ-MuLV *env* gene and the 3'-untranslated region of the *c-fos* gene. The 104-base pair sequence in the *c-fos* gene that is missing in the *v-fos* sequence is bounded by a 5-nucleotide inverted repeat which could have been looped out in the formation of FBJ-MuSV.

Despite their differences at the C-terminus, the *v-fos* and *c-fos* gene products are both located in the nucleus⁶. Furthermore, the altered C-terminus of the *v-fos* protein does not appear to be what determines its ability to transform fibroblasts *in vitro*, since the *c-fos* gene can be activated to transform fibroblasts without altering the coding region. Two manipulations are required to activate *c-fos* genes: a transcriptional enhancer element (contained in the LTRs) has to be linked to the gene and an interaction between the C-terminal coding and downstream non-coding sequences, which seems to inhibit *c-fos* expression at the level of translation or RNA maturation⁷, has to be disrupted.

Expression of the *c-fos* gene is developmentally regulated. During prenatal development, it is expressed at high levels only in the extra-embryonal tissues (amnion, chorion and placenta)^{8,9}. The *c-fos* protein has been found in the nucleus of normal mouse amnion cells⁶ in concentrations that are equivalent to those of *v-fos* protein in cells transformed by FBJ-MuSV. Here, then, is the enigma: although *v-fos* and *c-fos* proteins have different C-termini, they have the same subcellular location and both can transform fibroblasts *in vitro*; yet amnion cells expressing high levels of *c-fos* protein in their nucleus are not morphologically transformed. I submit that normal cellular *fos* protein can induce transformation when it is expressed in an inappropriate cell. □

Inder M. Verma is at the Molecular Biology and Virology Laboratory, The Salk Institute, San Diego, California 92138.

1. Finkel, M. P., Biskis, B. O. & Jenkins, P. B. *Science* **151**, 698 (1966).
2. Curran, T. *et al.* *J. Virol.* **44**, 674 (1982).
3. Van Beveren, C. *et al.* *Cell* **32**, 1241 (1983).
4. Curran, T. & Teich, N. M. *J. Virol.* **42**, 114 (1982).
5. van Straaten, F. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **80**, 3183 (1983).
6. Curran, T. *et al.* *Cell* **36**, 259 (1984).
7. Miller, A. D., Curran, T. & Verma, I. M. *Cell* **36**, 51 (1984).
8. Müller, R. *et al.* *Nature* **299**, 640 (1982).
9. Müller, R., Verma, I. M. & Adamson, E. A. *EMBO J.* **2**, 679 (1983).

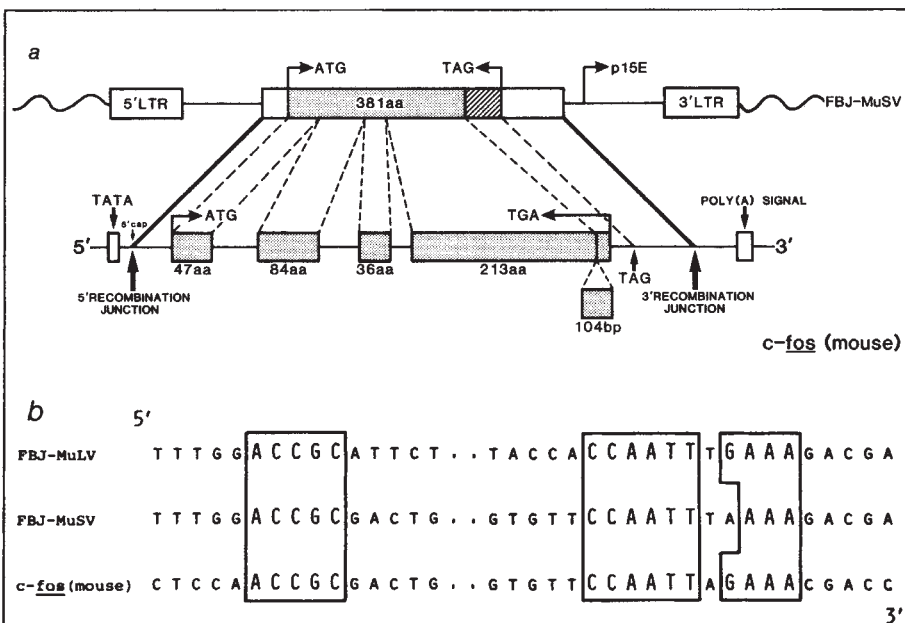


Fig. 1. *a*, Molecular architecture of FBJ-MuSV proviral DNA and mouse *c-fos* gene. FBJ-MuSV is shown at the top. The large open box indicates the acquired cellular sequences; solid bars, initiation and termination codons of *v-fos* proteins; the hatched box, the C-terminal 49 amino acids of *v-fos* protein encoded in a different reading frame because of deletion of 104 base pairs of *c-fos* sequence; bold lines, 5' and 3' recombination junctions. The position of p15E protein is indicated. In the *c-fos* gene below, the stippled boxes are the exons. The 104-base pair sequence that has been deleted in *v-fos* is indicated by a triangle. Unlike the *v-fos* protein, the *c-fos* protein terminates at a TGA codon. The position of the TAG codon which acts as chain terminator for *v-fos* protein is shown by an arrow. *b*, Nucleotide sequences at the 5' and 3' junctions of FBJ-MuLV, FBJ-MuSV and mouse *c-fos* gene. Shared sequences are boxed.

Multiple-organ autoimmunity

SIR — Autoimmune diseases affecting the thyroid, parathyroids, adrenals, stomach and pancreatic islets are genetically associated¹ and patients with a disease affecting one of these organs show an increased prevalence of diseases affecting others, as do their close relatives. This observation has given rise to the rather obvious suggestion that some patients may produce autoantibodies that recognize antigenic determinants which are common to two or more of these organs^{2,3}. Recent studies, reported in *Nature*, showing that virally-induced monoclonal antibodies in mice sometimes cross-react with several organs have given renewed prominence to this explanation.

We have carried out extensive absorption studies to determine whether the simultaneous occurrence of two or more auto-immune diseases in one patient can be explained simply by autoantibodies which cross-react with the different organs involved. Sera containing autoantibodies reactive with two autoantigens were absorbed with one of those antigens and then tested for residual activity against that antigen and also against the other. The combinations tested were: thyroid microsomal autoantibody (TMAab) and gastric parietal cell autoantibody (PCAab); thyroid stimulating autoantibody (TSAab) and TMAab; thyroglobulin autoantibody (TGAab) and TMAab. In none of these cases could we find the cross-reaction predicted; thus the autoantibody does not explain co-occurrence of these diseases.

This does not, however, indicate that cross-reactions between multiple tissue antigens never occur. On the contrary, there is striking evidence that some of the autoantibodies observed in systemic lupus erythematosus (SLE) recognize antigenic determinants occurring in a variety of seemingly unrelated molecules has provided a great simplification of a disease which was formerly bewildering⁴.

Further evidence against the simple cross-reaction explanation is the discordance in time of onset of multiple autoimmune diseases in individual patients¹. If cross-reactivity were the explanation, one might expect that in patients with two autoimmune diseases, both would appear simultaneously; this is not usually the case. Thus in only 10 per cent of 113 patients with both Addison's disease and insulin-dependent diabetes did both diseases present simultaneously; the average interval in the remainder was nearly 6 years⁵. Similarly, in a series of patients with pernicious anaemia and Graves' disease, the mean interval between diagnosis of the two conditions was 3.3 years with a range of 0–9 years⁶. These observations point to the likely role of somatic mutation in generating the immune diversity required for reaction with these diverse autoantigens and suggest

that the genetic influence must act on precursors of the pathogenic clones rather than on the clones themselves.

Advocates of "disordered immune regulation" notwithstanding, it seems obvious that specific autoimmune diseases must be a consequence of an unfortunate clonal specificity within an individual's immune response repertoire⁷. On this view, individuals not genetically predisposed to a particular disease will either have deleted the precursor of a particular forbidden clone or will possess anti-idiotypic clones capable of deleting the forbidden clones as they arise.

A corollary of this view is that in patients with multiple autoimmune diseases, the pathogenic clones are likely to have arisen by separate somatic mutations from a common precursor, or from separate precursors which share idiotypic determinants¹, thereby explaining why such diseases do not usually appear simultaneously. Sharing of idiotypic determinants between antibodies directed against closely related antigens seems most probable even though cross-reaction is not apparent in the paratopes of such antibodies¹. We predict that this is the key to understanding the occurrences of multiple autoimmune diseases in individual patients.

JOHN G. KNIGHT

ALLISON KNIGHT

Otago University Medical School,
PO Box 913,
Dunedin, New Zealand

1. Knight, J.G. & Adams, D.D. *Ciba Fdn Symp.* 90, 35–56.
2. Haspel, M.V. *et al. Nature* 304, 73–76 (1983).
3. Steinberg, A.D. *Nature* 304, 20 (1983).
4. Schwartz, R.S. *Immun. Today* 4, 68–69 (1983).
5. Solomon, N. *et al. Diabetes* 14, 300–304 (1965).
6. McNichol, G.P. *Am. J. med. Sci.* 241, 396–1961.
7. Knight, J.G. *Immun. Today* 3, 326–328 (1982).
8. Knight, J.G. & Knight, A. *J. clin. Immun.* (in the press).

Supernovae and life

SIR — Rubenstein *et al.*¹ have suggested that optical synchrotron emission from supernovae might be substantially circularly polarized. However, the optical emission from the Crab Nebula has < 1% circular polarization^{2,3}, and there are simple reasons for believing that naturally occurring synchrotron emission from extremely relativistic electrons will have a very low degree of circular polarization.

The synchrotron emission from an electron is confined within a cone, approximately aligned with its instantaneous direction of motion, and having a half-width of $\sim (mc^2)/E$ (rad). For electrons emitting in the visible, even if the field is as high as 10 G, $(mc^2)/E \approx 10^{-3}$. The emission from a single electron is 100% polarized — linearly polarized near the axis of the emission cone, RH elliptically polarized on the side of the cone towards the direction of magnetic field, and LH elliptically polar-

ized on the opposite side⁴. At the half-power points the degree of circular polarization is about 0.4.

The synchrotron emission from a group of relativistic electrons will have these same polarization properties only if the "pitch angles" of the spiral motions of the electrons all lie within a range which is less than the width of the emission cone of a single electron — that is, $< (mc^2)/E$. This condition is evidently satisfied for electrons in accelerators where considerable care is taken to ensure that the electron pitch angles are closely equal to 90°. It is unlikely to be true in supernovae.

When the range of electron pitch angles is large compared with the width of the individual emission cones, the radiation emitted in a particular direction comes almost equally from the RH and LH polarized sides of individual emission cones. The circularly polarized component is thus largely cancelled and the resulting degree of circular polarization becomes⁵ $0(mc^2)/E$.

In summary, if synchrotron emission is to be substantially circularly polarized the pitch angles of the electrons must be confined within a range less than the width of the emission cone of individual electrons: this includes the special case discussed by Epstein⁶. For extremely relativistic electrons the emission cone is so small that it is unlikely that this condition will be met in astronomical objects. Substantial circular polarization can arise when the electrons are barely relativistic and the emission is at low harmonics of the gyro-frequency⁷ ("gyro-synchrotron" emission). Since strong magnetic fields are believed to result when stars collapse, gyro-synchrotron emission might possibly provide Rubenstein *et al.* with a source of circularly polarized radiation in supernovae.

J. A. ROBERTS

Division of Radiophysics, CSIRO,
PO Box 76,
Epping, NSW 2121, Australia

1. Rubenstein, E., Bonner, W.A., Noyes, H.P. & Brown, G.S. *Nature* 306, 118 (1983).
2. Landstreet, J.D. & Angel, J.R.P. *Nature* 230, 103 (1971).
3. Martin, P.G., Illing, R. & Angel, J.R.P. *Mon. Not. R. astr. Soc.* 159, 191–201 (1972).
4. Korchak, A.A. & Syrovatskii, S.I. *Astr. Zh.* 38, 885–897 (1961); *Soviet Astr.-AJ* 5, 678–686 (1962).
5. Legg, M.P.C. & Westfold, K.C. *Astrophys. J.* 154, 499–514 (1968).
6. Epstein, R.I. *Astrophys. J.* 183, 593–610 (1973).
7. Channugam, G. & Dulk, G.A. *Astrophys. J.* 244, 569–578 (1981).

Scientific Correspondence

Scientific Correspondence is intended to provide a forum in which readers may raise points of a rather technical character which are not provoked by articles or letters previously published (where Matters Arising remains appropriate).

Thermodynamics and the end of a closed Universe

S. A. Bludman

Department of Physics, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA

If the Universe is closed and began with low entropy, it began without an initial singularity in an unstable de Sitter phase. Because of the irreversible nature of the processes, when this closed Universe ultimately contracts it will not bounce, but will go into a 'final crunch'. In this final reheating, symmetry will be restored, but the contraction will never become deflationary. A closed Universe therefore lives only once, and can start with a whimper, but goes out with a bang.

TWO most important conceptual (and ultimately practical) questions in cosmology are whether space is closed (positive curvature $k = +1$) or open (negative curvature $k = -1$ or flat $k = 0$) and whether the present expansion will continue monotonically or reverse itself some time in the future. In a universe with vanishing cosmological constant (Friedman universe), which may be the case in our Universe now, an open universe will expand monotonically and a closed universe will be ultimately reversing. Here, we begin by reviewing significant recent limits on the cosmological parameters λ_0 , q_0 , Ω_0 , H_0 , assuming only global homogeneity and isotropy, but not $\lambda_0 = 0$ or flatness. We show that, so far as present observations are concerned, the Universe may be open or closed, monotonically expanding or reversing. We then show (in agreement with Guth and Sher¹ but not with Petrosian²) that even if the present expansion were to reverse, the Universe cannot bounce in the future, but must end in a 'final crunch'. Our proof does not depend on particle physics, but follows directly from conventional thermodynamic stability requirements³ in any gravitational theory incorporating the newtonian limit and the equivalence and Birkhoff principles. Finally, we show that if space is closed and the Universe began with low entropy, then it had to begin, not with a big bang, but with a non-singular tepid little bang⁴.

Throughout this article, we assume global homogeneity and isotropy and conventional thermodynamics. We do not consider quantum gravity or gravitational entropy associated with matter clumping. The only gravitational law we assume is the Friedman equation

$$H^2 = \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G\tilde{\rho}}{3} - \frac{k}{a^2} \quad (1)$$

for the rate of expansion or contraction of a homogeneous isotropic universe. In this equation, $\tilde{\rho} = \rho + \rho_v$ is the total mass density of matter or radiation (ρ) plus vacuum (ρ_v) and $k = 0, \pm 1$ is the spatial curvature index. Equation (1) must hold in any gravitational theory that becomes newtonian in the non-relativistic limit and that incorporates the equivalence and Birkhoff principles^{5,6}. It is hard to imagine any fundamental gravitational theory, including a (future) theory of quantum gravity, in which equation (1) would not hold for a homogeneous isotropic cosmology.

Our thermodynamic assumptions are as conservative as our gravitational assumptions. Following Einstein⁷, we assume that classical thermodynamics "is the only physical theory of universal content concerning which I am convinced that within the framework of applicability of its basic concepts, it will never be overthrown". For the energy and pressure in co-moving volume $V \sim a^3$, we assume $dE + \tilde{p}dV = 0$, or in terms of energy density $E/V \equiv \tilde{\epsilon}$, $d\tilde{\epsilon} = -(\tilde{\epsilon} + \tilde{p})dV/V$. From equation (1) we then obtain

$$\ddot{a} = -\frac{4\pi G}{3}(\tilde{\rho} + 3\tilde{p}c^{-2})a \quad (2)$$

the relativistic term following simply from $\tilde{\epsilon} = \tilde{\rho}c^2$. Since we treat only a homogeneous isotropic universe (vanishing Weyl tensor), we ignore gravitational clumping or black hole formation or evaporation: we do not introduce any 'gravitational entropy'⁸. Because we assume conventional thermodynamics, the ordinary entropy cannot decrease. Since the expanding or contracting universe is not in true thermal equilibrium, the usual thermodynamic inequalities need not apply. Nevertheless, although temperature and other thermodynamic quantities are strictly defined only near thermal equilibrium, we do deal with metastable states and calculate the changes in entropy between states that are assumed to be nearly in equilibrium.

Our assumption of conventional thermodynamics requires the approach to equilibrium as the universal expansion slows down and the material density decreases. This assumption of asymptotic equilibrium is therefore equivalent to the vanishing of the cosmological constant in the present cold, low-density phase. Classical thermodynamic stability therefore requires that a contracting universe does not bounce and that the present cosmological constant vanishes.

The present Universe

In this section we review the observational data on maximum redshift, acceleration, matter density, Hubble constant and age of the present Universe. This review is motivated by the improved limits for Ω_0 and t_0 and by the predictions of inflationary theory. We assume only a homogeneous isotropic Universe, but do not assume vanishing cosmological constant, open space or a big bang. We find that the observations are consistent with, but do not require, $\lambda_0 \approx 0$ and $\Omega_0 \approx 1$. In any case, space may be open or closed and the present expansion may continue monotonically or be reversed.

From equations (1) and (2), the scale-invariant matter density $\Omega = 2\sigma = \tilde{\rho}/\rho_{CR} = 8\pi G\tilde{\rho}/3H^2$, vacuum density $\lambda = \rho_v/\rho_{CR}$ and deceleration have the present values (subscript 0)

$$1 = \Omega_0 + \lambda_0 - k/\dot{a}_0^2 \quad (3)$$

$$q_0 = \frac{1}{2}\Omega_0 - \lambda_0 = \sigma_0 - \lambda_0 \quad (4)$$

The factor $\frac{1}{2}$ arises because, in the present Universe, we are neglecting radiation pressure so that $\rho = \rho_0(a_0/a)^3$. In terms of redshift $1+z = a_0/a$, these quantities evolve as

$$\rho_{CR} = \rho + \rho_v + (\rho_{CRO} - \rho - \rho_v)(1+z)^2$$

$$(H/H_0)^2 = \Omega_0(1+z)^3 + \lambda_0 + (1 - \Omega_0 - \lambda_0)(1+z)^2 \quad (5)$$

$$\Omega/\Omega_0 = (1+z)^3/[\Omega_0(1+z)^3 + \lambda_0 + (1 - \Omega_0 - \lambda_0)(1+z)^2] \quad (6)$$

so that $\Omega \rightarrow 1$ in the distant past and $\rightarrow 0$ in the far future. If the near flatness of the present Universe is to be insensitive to the present epoch, $\lambda_0 \approx 0$, $\Omega_0 \approx 1$ is necessary.

Because the Universe has expanded and cooled at least since nucleosynthesis at $a/a_0 = x < 10^{-10}$, $H^2 \propto \sigma_0 - q_0$ $x^3 + (1 - 3\sigma_0 - q_0)x + 2\sigma_0$ must remain positive for $0 < x \leq 1$. For $k \leq 0$ ($1 - 3\sigma_0 - q_0 \geq 0$) this requires $\sigma_0 > q_0$, but for $k > 0$ ($1 - 3\sigma_0 - q_0 < 0$) this requires $\sigma_0 > \sigma_E$ where

$$27(\sigma_E - q_0)\sigma_E^2 = (3\sigma_E - q_0 - 1)^3 \quad (7)$$

or

$$\sigma_E = \frac{(q_0 + 1)}{6} [q_0 + 1 + \sqrt{(q_0 + 1)(q_0 - 1/3)}] \quad (8)$$

Conversely, for each σ_0 the right-hand side of equation (8) furnishes a lower bound $q_E(\sigma_0)$ and an upper bound $\lambda_E(\sigma_0)$ on q_0, λ . If $\sigma_0 < 0.06, 0.12, 0.42, 1.53$, the present acceleration and cosmological constant cannot exceed $-q_E = 1.33, 1.5, 2, 3, \lambda_E = 1.4, 1.6, 2.4, 4.5$, respectively. For any reasonable value of the present matter density, the present cosmological constant, even if not zero, is small enough to have negligible effect on gravitational dynamics on all but the largest cosmological scales.

Integrating equation (5)

$$H_0 t_0 = \int_{x_m}^1 \frac{dx}{x \sqrt{\Omega_0 x^{-1} + (1 - \Omega_0 - \lambda_0) + \lambda_0 x^2}} \quad (9)$$

is the time since the Universe expanded from minimum scale $a_0 x_m$. The requirement that the Universe has expanded at least by 10^{10} makes x_m negligibly small or zero. With small corrections for the vacuum- and radiation-dominated epochs preceding the current matter-dominated epoch, t_0 is practically the age of the Universe since the big bang or since the last bounce. Figure 1 shows $H_0 t_0$ as function of $\log \sigma_0$ for fixed q_0 . Along the dashed line $k = 0$ ($\lambda_0 + \Omega_0 = 1$)

$$H_0 t_0 = g(\Omega_0) = \frac{2}{3} \lambda_0^{-1/2} \tanh^{-1} \lambda_0^{1/2} \\ = \frac{2}{3} (1 - \Omega_0)^{-1/2} \operatorname{sech}^{-1} \Omega_0^{1/2} \quad (10)$$

$H_0 t_0 \rightarrow 2/3$ for $\lambda_0 = 0$ and $\rightarrow 1/3 (\ln 2 - \ln \sigma_0)$ for $\Omega_0 = 2\sigma_0 \ll 1$.

The heavy line in Fig. 1 separates universes which (because of positive space curvature or negative cosmological constant in equation (5)) will ultimately reverse their expansion from those which (because of negative space curvature or positive cosmological constant) expand monotonically. The equation of this heavy line is $\lambda_0 = \lambda_E(\sigma_0)$ for $\Omega_0 > 1$ ($H_0 t_0 < 2/3$) and $\lambda_0 = 0$ for $\Omega_0 < 1$ ($H_0 t_0 > 2/3$) in which case

$$H_0 t_0 = f(\sigma_0) \equiv (1 - 2\sigma_0)^{-1} \\ - \sigma_0 (1 - 2\sigma_0)^{-3/2} \cosh^{-1}(\sigma_0^{-1} - 1) < 1 \quad (11)$$

The curves for zero curvature (10) and zero cosmological constant (11) cross at $\Omega_0 = 1$, $H_0 t_0 = 2/3$. If $H_0 t_0 > 2/3$, a closed space must expand monotonically, but an open space expands monotonically for $\lambda_0 > 0$, and reverses for $\lambda_0 < 0$: if $\Omega_0 > 1$, this requires that the universe be closed, nevertheless monotonically expanding; if $\Omega_0 < 1$, the expansion will be reversing for $2/3 < H_0 t_0 < f(\sigma_0)$, monotonic for $f(\sigma_0) < H_0 t_0$.

Because of equations (4) and (9), a homogeneous isotropic cosmology is determined by any two of the four dimensionless parameters $\Omega_0, \lambda_0, q_0, H_0 t_0$. Of these, the deceleration q_0 is least determined and the dimensionless age $H_0 t_0$ is best determined. Years of study of the Hubble diagram show^{9,10,22} that the apparent deceleration is $> 1/2$ so that, with reasonable corrections for dynamic and stellar evolution, the true acceleration is probably $q_0 > -1.0$. This equivocal evidence does no more than suggest that any cosmological constant $|\lambda_0|$ cannot be large compared with σ_0 and that the Universe is at least approximately flat.

For $\lambda_0 \approx 0$, the dynamical evidence on Ω_0 from the virgocentric flow suggests¹¹ $\Omega_0 = 0.2-0.5$, consistent with the value $\Omega_0 = 0.07-0.6$ from the cosmic virial theorem¹². If the virgocentric infall velocity has been underestimated, if $\lambda_0 = 0$, if neutrinos or other light mass particles are not clumped in clusters of galaxies,

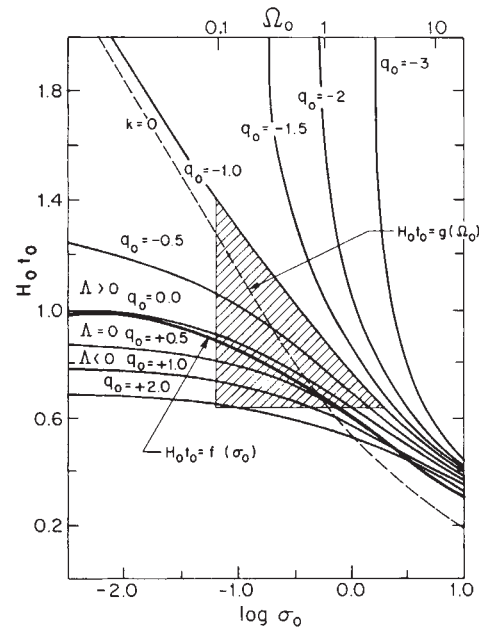


Fig. 1 Dimensionless age of the present Universe as function of dimensionless present density $\Omega_0 = 2\sigma_0$ and deceleration $q_0 = \sigma_0 - \lambda_0$. Below (above) the dashed curve $H_0 t_0 = g(\Omega_0)$ the Universe is open (closed). Below (above) the heavy solid curve $H_0 t_0 = f(\sigma_0)$ the expansion is reversing (monotonic). The shaded region is that allowed by $\Omega_0 > 0.1$, $q_0 > -1.0$ and $t_0 > 12$ Gyr.

or if these clusters are not virialized, these dynamical values for Ω_0 are only lower limits. The limits $q_0 < -1.0$, $\Omega_0 > 0.1$ derived from the Hubble diagram and dynamical measurements define the two upper boundaries for the shaded region in Fig. 1. $H_0 t_0 < 1.5$ if $q_0 > -1$ and indeed < 1.1 if $q_0 > -0.5$.

The age of globular clusters ($t_0 = 18 \pm 2$ Gyr (ref. 13), 15 ± 3 (ref. 14) or the nucleochronological age of the Galaxy¹⁵ ($t_0 = 20.8^{+2.4}_{-2.2}$ Gyr) would now determine the cosmological model if only the systematic discrepancy in the reported values $H_0 = 95 \pm 5$ (refs 16, 17), 82 ± 10 (ref. 18), 52 ± 6 (ref. 19) $\text{km s}^{-1} \text{Mpc}^{-1}$ for the Hubble constant could be resolved. These measurements of t_0, H_0 define the lower boundary $0.63 < H_0 t_0$ in Fig. 1. Although $H_0 t_0 = 1.17 \pm 0.3$ would be consistent with all measurements, the quoted errors make the measured values of H_0 mutually exclusive. We therefore discuss separately the implications of the two extreme values for H_0 .

If $H_0 = 95 \pm 5 \text{ km s}^{-1} \text{Mpc}^{-1}$ (refs 16, 17) ($H_0^{-1} = 10.3 \pm 0.5$ Gyr) so that $H_0 t_0 = 1.6 \pm 0.5$, then the Universe must be presently vacuum-dominated ($\lambda_0 \gg \Omega_0 \sim 0.1$) and accelerating ($-q_0 \approx \lambda_0 > 1/2$); although space may be closed, the present expansion will continue monotonically. This large value of H_0 is inconsistent with $q_0 > -0.5$ or $\Omega_0 > 0.3$ and with inflationary theory which requires $\lambda_0 \approx 0$, $\Omega_0 \approx 1$.

On the other hand, if $H_0 = 52 \pm 6 \text{ km s}^{-1} \text{Mpc}^{-1}$ (ref. 19) ($H_0^{-1} = 18.8 \pm 2.2$ Gyr) so that $H_0 t_0 = 0.88 \pm 0.25$, then $0.2 < \Omega_0 < 2$ and only a reversing universe of negative curvature is excluded: if $q_0 \sim 0.5$, then $\Omega_0 \approx 0.1$ and $\lambda_0 < 0$ so that, although space is open, the present expansion will reverse itself; if $q_0 < 0.3$ or $\Omega_0 > 0.2$, then $\lambda_0 > 0$ and the present expansion will continue monotonically, whether space is open or closed. The lower value of H_0 allows flat space and allows $\lambda_0 = 0$ if $\Omega_0 \approx 0.1$ or if $\Omega_0 = 1$ and $t_0 = \frac{2}{3} H_0^{-1} = 12.5 \pm 1.4$ Gyr.

For either value of H_0 , if $\Omega_0 > 0.2$ a closed space is possible. Indeed, space must be closed if $\Omega_0 > 1$ and/or $H_0 t_0 > 1.3$.

While an age $t_0 = 11.5 \pm 0.5$ Gyr would be consistent with flat space for either value of H_0 , the lower value of H_0 then requires $\Omega_0 \approx 1$, $\lambda_0 = 0$ and the larger value requires $\Omega_0 = 0.2$, $\lambda_0 = 0.8$.

The inflationary scenario²⁰ builds on the premise that the near flatness of the present Universe is no temporary accident;

for the matter plus vacuum energy densities to remain close to ρ_{CR} over a broad range of epochs requires that $\lambda_0 = 0$, $\Omega_0 \approx 1$, $H_0 t_0 = 2/3$. Most inflationary theories assume not only $\lambda_0 = 0$ exactly but also that the Universe started with a big bang. These two assumptions are so far as much articles of faith as once were Einstein's contrary assumptions of a cosmological constant and closed space. Nevertheless, the remainder of this article assumes that at present $\lambda_0 = 0$ and $\Omega_0 \approx 1$, but considers the possibility that Ω_0 slightly exceeds unity so that space is closed.

Expanding and contracting universes

We now consider the dominating gravitational effects of the huge cosmological constant induced by elementary particle symmetry restoration in the very early and very late universes.

In the present Universe, elementary particle symmetry is hidden by the presence of several order parameters σ which 'break' the symmetry. In the early hot Universe ($T > T_c$), these order parameters vanished so that the present hidden symmetry was unbroken. As the Universe expanded and cooled below $T_c \sim \sigma$, the Lorentz-invariant vacuum (elementary particle ground state) underwent a series of phase transitions, each with the release of a huge latent heat density $L \sim m_Y^4$, where m_Y is the mass acquired by the gauge bosons in this original first-order phase transition. For grand unification theories (GUTs) $\sigma \sim 5 \times 10^{14}$ GeV, but if there was an earlier phase transition associated with gravity, $\sigma_Y \sim m_{PL} = 1.2 \times 10^{19}$ GeV.

Just because the energy density of empty space (cosmological constant) is observed to be small or zero at present, it must have been huge in the very early and very late Universes, when its gravitational effects can be most important. In any particle theory in which symmetry is restored in the high-temperature phase, there must be a huge cosmological constant induced in the very early Universe and, if the expansion reverses, in the very late Universe.

In order that the stress-energy

$$T^{\mu\nu} = (\tilde{\epsilon} + \tilde{p})u^\mu u^\nu - \tilde{p}g^{\mu\nu}$$

be, for the vacuum, independent of frame velocity u^μ , we must require at all temperatures $(\tilde{\epsilon} + \tilde{p})_v = 0$. Here $g^{\mu\nu}$ is the metric, $(-1, -1, -1, 1)$ in Minkowski space. In the early Universe the matter density is negligible, so that the energy density and pressure can be projected into vacuum and radiation parts

$$\tilde{\epsilon} = \frac{1}{4}(\tilde{\epsilon} - 3\tilde{p}) + \frac{3}{4}(\tilde{\epsilon} + \tilde{p}) \equiv \epsilon_v + \epsilon \quad (12)$$

$$\tilde{p} = -\frac{1}{4}(\tilde{\epsilon} - 3\tilde{p}) + \frac{1}{4}(\tilde{\epsilon} + \tilde{p}) \equiv p_v + p \quad (13)$$

In the vacuum sector, all chemical potentials vanish, the free energy density $f = (E - TS)/V = \tilde{\epsilon} - Ts = -\tilde{p}$, all pressures and energy densities are functions of temperature T only, and the energy density $s = -df/dT$. At every temperature, $\epsilon(T) = 3p(T) = \frac{1}{4}Ts$, $\epsilon_v(T) = -p_v(T) = \frac{1}{4}Ts + f$. The vacuum energy dominates the radiation energy and $\tilde{\epsilon} + 3\tilde{p} < 0$ whenever the temperature falls below T_v defined by $\epsilon_v(T_v) = \epsilon(T_v)$. In the vacuum dominated phase the total pressure $\tilde{p} < -2p < 0$.

At finite temperature the effective Higgs potential is

$$V_{\text{eff}}(\phi, T) = V_{\text{eff}}^0(\phi) - G(\phi, T) \quad (14)$$

where $V_{\text{eff}}(\phi)$ is the zero-temperature effective potential and $G(\phi, T)$ is a homogeneous function of fourth degree in ϕ , T . The order parameter $\sigma(T)$ is determined implicitly by

$$\left(\frac{\partial V_{\text{eff}}}{\partial \phi} \right)_{\phi=\sigma(T)} = 0 \quad (15)$$

and remains practically equal to its zero-temperature value σ_0 , up to a temperature T_2 at which it rapidly goes to zero.

The free energy density, $f(T) = V_{\text{eff}}(\sigma(T), T)$ and the entropy density $s = d\tilde{p}/dT = \partial G/\partial T$, so that

$$Ts = 4G - \phi \left(\frac{\partial G}{\partial \phi} \right)_\sigma = 4G - \sigma \frac{\partial V_{\text{eff}}^0}{\partial \sigma}$$

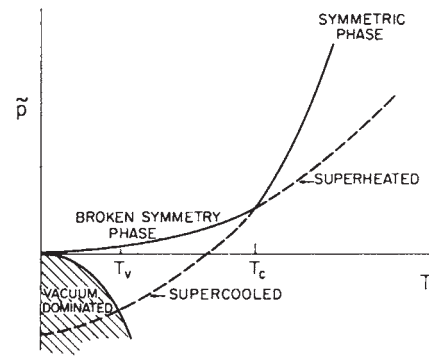


Fig. 2 In the vacuum sector the total pressure is a convex function of temperature which, because of the vanishing cosmological constant, vanishes at $T = 0$. The discontinuity in $s = d\tilde{p}/dT$ at temperature T_c indicates a first-order phase transition. The heavy line shows the stable phase, the dashed lines the metastable phases. A vacuum-dominated Universe must have negative pressures.

The radiation pressure $p = \partial(\phi^{-4}G)/\partial\phi^{-4} \approx G$, except during the phase transition. The vacuum energy density

$$\epsilon_v(T) = -p_v(T) = v_{\text{eff}}^0 - \frac{1}{4}\sigma \frac{\partial V_{\text{eff}}^0}{\partial \sigma} = \frac{d}{d\sigma^{-4}} (\sigma^{-4} V_{\text{eff}}^0)$$

depends only on $V_{\text{eff}}^0(\sigma)$, specifically on the part that is not simply quartic in ϕ . Petrosian² assumes the Coleman-Weinberg form

$$V_{\text{eff}}^0(\phi) = \epsilon_v(0) \left[1 - \left(\frac{\phi}{\sigma_0} \right)^4 \left(1 - 4 \ln \frac{\phi}{\sigma_0} \right) \right]$$

and obtains $\epsilon_v = \epsilon_v(0) [1 - (\sigma/\sigma_0)^4]$.

During the phase transition, the vacuum energy density drops rapidly from one constant value $\epsilon_v(0)$ to another constant value 0. This change in vacuum energy is determined by energy conservation $d\epsilon = T ds = \frac{4}{3}T d(\epsilon/T)$, so that $d\epsilon_v = \frac{1}{3}\epsilon d \ln(\epsilon T^{-4})$. If $g(T)$, the effective number of massless degrees of freedom, is defined by $\epsilon = (\pi^2/15)g(T)T^4$, then $d\epsilon_v = p d \ln g$, showing that the rapid decrease in vacuum energy is brought about by the decrease in the number of massless degrees of freedom during the phase transition.

The stable phase is always the phase of minimum free energy or maximum pressure $\tilde{p} = -f$. In thermal equilibrium, the pressure $\tilde{p} = p_v + p$ must be positive^{1,3} so that the vacuum-dominated de Sitter phase can be, at most, a supercooled state of metastable equilibrium.

The free energy, and in our case the free energy density, is a concave function of T : the entropy density is an increasing function of T . Together with the vanishing cosmological constant in the low-temperature phase, this makes $\tilde{p}(T)$ a convex function of T vanishing at $T = 0$ (Fig. 2). This lets the supercooled phase become vacuum-dominated as soon as T falls below T_v , but the metastable superheated phase, because it has $\tilde{p} > 0$, must always be radiation or matter dominated.

Thus, in any contraction leading towards equilibrium, there can be no vacuum dominance (deflation)²¹. Even if the contraction were to lead far away from equilibrium², it seems likely that such a strongly irreversible process will raise the entropy Σ' to such a high value that there will be no bounce. Unless new laws (quantum gravity) intervene, a closed universe on contraction undergoes an 'ultimate crunch'.

In fact, the metastable state reached by rapid expansion can inflate and supercool a very great deal before snapping back to equilibrium. The metastable broken-symmetry phase reached by superheating on rapid contraction becomes, on the contrary, suddenly unstable when the temperature heats up to T_2 at which $\sigma(T)$ vanishes. Protracted supercooling generates a great deal

of entropy, but the entropy generated on contraction is limited, if equilibrium is ever re-established.

Some authors speculate about new gravitational effects in the Planck era allowing a closed universe to bounce repeatedly. We would argue that in any reasonable gravitational theory incorporating the equivalence and Birkhoff principles, the dynamical contraction of a homogeneous isotropic universe will still be governed by an equation like equation (1) so that our Universe cannot bounce in the future. Closed Friedman universes were once called oscillatory universes. We now appreciate that, because of the huge entropy generated in our Universe, far from oscillating, a closed universe can only go through one cycle of expansion and contraction. Whether closed or open, reversing or monotonically expanding, the severely irreversible phase transitions transpiring give the Universe a definite beginning, middle and end.

A non-singular origin?

Before the first symmetry-breaking phase transition, the Universe was dominated by the vacuum energy density $\varepsilon_v(0)$ and the radiation energy density $\varepsilon = a_s' T^4 = \frac{1}{15} \pi^2 g' T^4$, where the effective number of massless degrees of freedom $g' > 100$. The conserved total entropy $\Sigma' = \frac{4}{3} a_s' (aT)^3$ defines a length scale

$$a_R'^2 = \frac{8\pi G}{3} \varepsilon a^4 = 2\pi (\frac{3}{4} a_s')^{1/3} \Sigma'^{4/3} m_{PL}^{-2}$$

and the vacuum energy a second length scale

$$b^{-2} = \frac{8\pi G}{3} \varepsilon_v \sim m_Y^4 m_{PL}^{-2}$$

Since $m_Y^{-1} = 10^{-31 \pm 2}$ cm, the mean of these two scales

$$a_E = (a_R' b)^{1/2} = (\varepsilon a^4 / \varepsilon_v)^{1/4} \sim a_s'^{-1/2} \Sigma'^{1/3} m_Y^{-1} \quad (16)$$

is a very small scale at which $\varepsilon = \varepsilon_v$ and $T = T_v$. The ratio of these two scales

$$\alpha = \left(\frac{2a_R'}{b} \right)^2 = \frac{64\pi^2}{3} \left(\frac{3}{4a_s'} \right)^{1/3} \Sigma'^{4/3} \frac{\varepsilon_v}{m_{PL}^4} \sim 5a_s'^{-1/3} \Sigma'^{4/3} (m_Y / m_{PL})^4 \quad (17)$$

is a dimensionless measure of the radiation energy or total entropy Σ' in the symmetric universe before the first phase transition. Along this initial adiabat, $aT = \text{constant} = a_E T_v$. Because almost all the entropy multiplication to the present value $\Sigma_0 \gg 10^{87}$ took place during the drawn-out inflationary epoch of the later GUTs, we expect the initial entropy Σ' and α to be small. The behaviour of the very early Universe depends critically on the spatial curvature k and on α .

The Friedman equations

$$\dot{a}^2 = \frac{8\pi G}{3} \varepsilon a^2 - k = \left(\frac{a}{b} \right)^2 + \left(\frac{a_R'}{a} \right)^2 - k \quad (18)$$

$$\ddot{a} = -\frac{4\pi G}{3} (\varepsilon + 3pc^{-2}) = \frac{a}{b^2} - \frac{a_R'^2}{a^3} \quad (19)$$

show that there is always an inflection at $a = a_E$ when, at temperature $T = T_v$, we pass between radiation and vacuum energy dominance; $\dot{a}$ can vanish only if both $k = +1$ (spherical universe) and $\alpha \leq 1$ (low entropy). Thus, for homogeneous isotropic radiation-dominated universes with positive vacuum energy, there are two possibilities.

(1) If the Universe is open ($k = -1$) and/or $\alpha > 1$ (relatively high entropy initial state), there can be no extreme in a . We have an inflectional Universe with a singular origin (hot big bang) at which $\varepsilon \gg \varepsilon_v$. This inflectional model in which the curvature term was insignificant in the early Universe, is the one advocates of the inflationary scenario usually have in mind.

The present Universe already has entropy $\Sigma_0 > 10^{87}$. Even if it is closed and undergoes recontraction, reheating and symmetry restoration, α will then be $\gg 1$, so that as we saw in the preceding section the ultimate crunch can never be reversed. Nor could it have bounced in the past if it began hot or developed a great deal of entropy in a first contraction, so that $\alpha > 1$.

(2) If the Universe is closed ($k = +1$) and $\alpha = \sin^2 2\phi < 1$ (low entropy initial state), then we might envisage a bouncing Universe where $\dot{a} = 0$ at either

$$a_{\max} = a_E \tan^{1/2} \phi, \quad T_{\min} = T_v \cot^{1/2} \phi$$

or at

$$a_{\min} = a_E \cot^{1/2} \phi, \quad T_{\max} = T_v \tan^{1/2} \phi$$

If there were a singular origin, the Universe would have remained very small ($0 \leq a \leq a_{\max} \leq a_E$) and very hot ($T \geq T_{\min} \geq T_v$), never becoming our present Universe. But it could have begun with $a \geq a_{\min} \geq a_E$, supercooled and vacuum dominated $T \leq T_{\max} \leq T_v \leq T_c$. This unstable initial Universe, which we call the tepid little bang⁴, would in the course of time undergo the first-order phase transitions to the present Universe.

Thus, if the Universe is open or began with significant entropy ($\alpha > 1$), it started in a big bang and later expanded into the unstable de Sitter phase. If the Universe is closed and (as seems reasonable) began with low entropy, it started non-singularly but already in the unstable de Sitter phase. In both cases, the subsequent GUTs inflation, baryosynthesis, nucleosynthesis and galaxy formation are the same. If the Universe is presently indeed nearly flat (as demanded by the inflationary scenario), then until the present expansion does or does not approach reversal, it will be practically impossible to distinguish between open big bang and closed tepid little bang universes. Until then, however, for those who prefer to maintain conventional gravitational and thermodynamic concepts, the tepid little bang enjoys the theoretical advantage of avoiding the initial singularity which would otherwise signal the incompleteness of Einstein's theory.

I thank Alan Guth, Marc Sher and Gary Steigman for helpful discussions. This research was supported in part by the US Department of Energy under contract EY-76-C-02-3071 and partly done at the Aspen Center for Physics.

Received 11 October 1983; accepted 5 January 1984.

- Guth, A. H. & Sher, M. *Nature* **302**, 505 (1983).
- Petrosian, V. *Nature* **302**, 506 (1983).
- Landau, L. D. & Lifshitz, L. M. *Statistical Physics*, 41, 262 (Pergamon, London, 1958).
- Bludman, S. A. *Proc. IAU Symp.* No. 104, 447-451 (1983).
- McCrea, W. H. & Milne, E. A. *Q. J. Math.* **5**, 73 (1934).
- Milne, E. A. *Q. J. Math.* **5**, 64 (1934).
- Einstein, A. in *Albert Einstein Philosopher-Scientist* (ed. Schelp, P. A.) 31 (Tudor, New York, 1949).
- Barrow, J. D. & Tipler, F. J. *Nature* **276**, 453 (1978).
- Fang, L. Z., Kiang, T., Cheng, F. H. & Hu, F. X. *Q. J. R. astr. Soc.* **23**, 363 (1982).
- Hoessel, J. G., Gunn, J. E. & Thuan, T. X. *Astrophys. J.* **241**, 486 (1980).

- Davis, M. & Peebles, P. J. E. *A. Rev. Astr. Astrophys.* **21**, 109 (1983).
- Press, W. H. & Davis, M. *Astrophys. J.* **259**, 449 (1981).
- Sandage, A. R. *Astr. J.* **88**, 1159 (1983).
- Iben, I. & Renzini, A. *Phys. Rep.* (in the press).
- Thielemann, F.-K., Metzinger, J. & Klapdor, H. V. *Astr. Astrophys.* **123**, 162 (1983).
- de Vaucouleurs, G., Peters, W. L., Bottinelli, L., Gauguenheim, L. & Paturel, G. *Astrophys. J.* **248**, 408, 86-B3 (1981).
- Aaronson, M. *et al. Astrophys. J.* **239**, 12 (1980).
- Aaronson, M. & Mould, J. *Astrophys. J.* **265**, 1 (1983).
- Sandage, A. R. & Tammann, G. A. *Astrophys. J.* **256**, 339 (1982).
- Guth, A. H. *Phys. Rev. D* **23**, 347 (1981).
- Davies, P. C. W. *Nature* **301**, 398 (1983).
- Hoessel, J. G. *Astrophys. J.* **241**, 493 (1980).

Structure of chromosomal rearrangements induced by the FB transposable element in *Drosophila*

Mary Collins* & Gerald M. Rubin*

Department of Embryology, Carnegie Institution of Washington, 115 West University Parkway, Baltimore, Maryland 21210, USA

Examination of 10 chromosomal rearrangements induced by foldback transposable elements in Drosophila demonstrates that sequences located between two closely linked elements are frequently deleted. Rearrangements and recombination events occurring within or adjacent to a foldback element are also observed. The type and frequency of rearrangement generated depend on element structure and the organization of flanking DNA.

THE rearrangement of DNA sequences is an important source of variability in eukaryotic genome evolution. Recent studies in prokaryotes, yeast, maize and *Drosophila* have implicated transposable elements in the induction of chromosomal rearrangements¹. We were interested in characterizing the structure of transposable element-induced rearrangements in a higher eukaryote in order to analyse the structure of rearrangement junctions, to determine which types of rearrangements are preferentially induced, and to learn more about rearrangement mechanisms.

The foldback (FB) family of transposable elements in *Drosophila* is capable of inducing chromosomal rearrangements at high frequencies. At least three independent cases of genetic instability described in *Drosophila* are caused by FB elements, suggesting the importance of these elements in generating new sequence arrangements^{2,3}. FB elements are dispersed, middle repetitive DNA sequences with long inverted repeats at their termini^{4,5}. These inverted repeats vary in length, and are composed of many small, tandemly arranged imperfect direct repeats⁶. The presence of long inverted repeats distinguishes FB elements from other characterized *Drosophila* transposable elements, Ty1 in yeast, and retroviruses¹.

We have used the *white-crimson* (w^c) mutation of *Drosophila* to screen for phenotypically detectable chromosomal rearrangements induced by an FB transposable element. This genetically unstable mutation is a recessive allele of the X-linked *white* eye colour locus, and was isolated as a partial phenotypic revertant of the mutant *white-ivory* (w^i) allele⁷. The w^i mutation results from the duplication of 2.9 kilobases (kb) of DNA within the *white* locus⁸. The w^c mutation results from the insertion of a 10-kb FB element into the w^i duplication (Fig. 1)^{2,9}. This FB element insertion is responsible for the mutability and darker eye colour of the w^c allele⁹.

w^c mutates at a frequency of > 1 in 10^3 X chromosomes to novel phenotypes, of which the most frequent classes are wild-type, white-ivory and white-eyed derivatives. Reversion of w^c to a wild-type phenotype occurs by an apparent recombination between the two flanking copies of the w^i duplication, while reversion to white-ivory is mediated by precise excision of the FB element^{9,10}. Cytological analysis of white-eyed derivatives indicates that at least some of these contain large deletions flanking the *white* locus⁷.

We have examined the structure of 10 white-eyed derivatives of w^c in order to elucidate chromosomal rearrangements induced by FB transposable elements. All 10 mutations were independently derived and selected for male viability and fer-

tility. This selection eliminates any large deletions, since there are essential loci closely linked to *white*¹¹.

Deletions

We screened all 10 chromosomes for alterations in *white* locus sequences by hybridizing genomic DNA blots with probes derived from the wild-type *white* locus, and found that 6 of the 10 chromosomes contained deletions. In all six cases, sequences to the right (centromeric) side of the insertion are conserved, while sequences directly adjacent to the w^c insertion and extending to the left (telomeric) side of the insertion are deleted (Figs 1, 2).

In all six cases the left deletion break point is contained within a 3.5-kb *Bam*HI-*Pvu*II fragment 14 kb to the left of the w^c insertion (Fig. 1). We examined this break-point region further to determine if a second FB element was present in the w^c chromosome. This region was isolated from a w^c background in a recombinant phage, and compared with the same region isolated from a wild-type chromosome¹². Fragments containing the break-point region from the w^c background are ~2 kb larger than homologous fragments from the wild-type chromosome. Furthermore, fragments from the break-point region hybridize to an FB element probe only when isolated from a w^c background; homologous fragments from the wild-type chromosome do not hybridize (Fig. 3). This indicates that the w^c chromosome contains an additional FB element insertion ~14 kb to the left of the w^c insertion. Homologous recombination between the w^c insertion and this closely linked FB element is the most likely mechanism for the generation of our six deletion chromosomes.

Recombination between two FB elements could occur at any of the homologous directly repeated sequences within the inverted repeats, leaving behind new FB elements of various lengths. The w^c FB element insertion has inverted repeats 2.2 and 3.4 kb in length, which are separated by a 4.0-kb w^c 'middle', a low copy number sequence which is not homologous to the inverted repeats (Fig. 1)². Recombination events involving the right inverted repeat of the w^c insertion should delete sequences from the w^c middle, while those involving the left inverted repeat should leave these sequences intact. We have detected both types of rearrangement (Fig. 1). Size estimates of the new FB elements generated by these deletions were made by measuring the distance between restriction sites which flank the residual elements; the distance from the *Pvu*II site marking the left deletion break point to the conserved *Xba*I site on the right of the w^c insertion varies from 3.5 to 10.5 kb in these six deletion derivatives (Fig. 2). These sizes are within the range predicted by a recombination mechanism. We have directly demonstrated that FB element DNA remains at the deletion break point for one derivative; the break-point region of this derivative isolated in a recombinant plasmid hybridizes to an FB element probe (Fig. 3).

* Present addresses: Genetics Institute, 225 Longwood Avenue, Boston, Massachusetts 02115, USA (M.C.); Department of Biochemistry, University of California at Berkeley, Berkeley, California 94720, USA (G.M.R.).

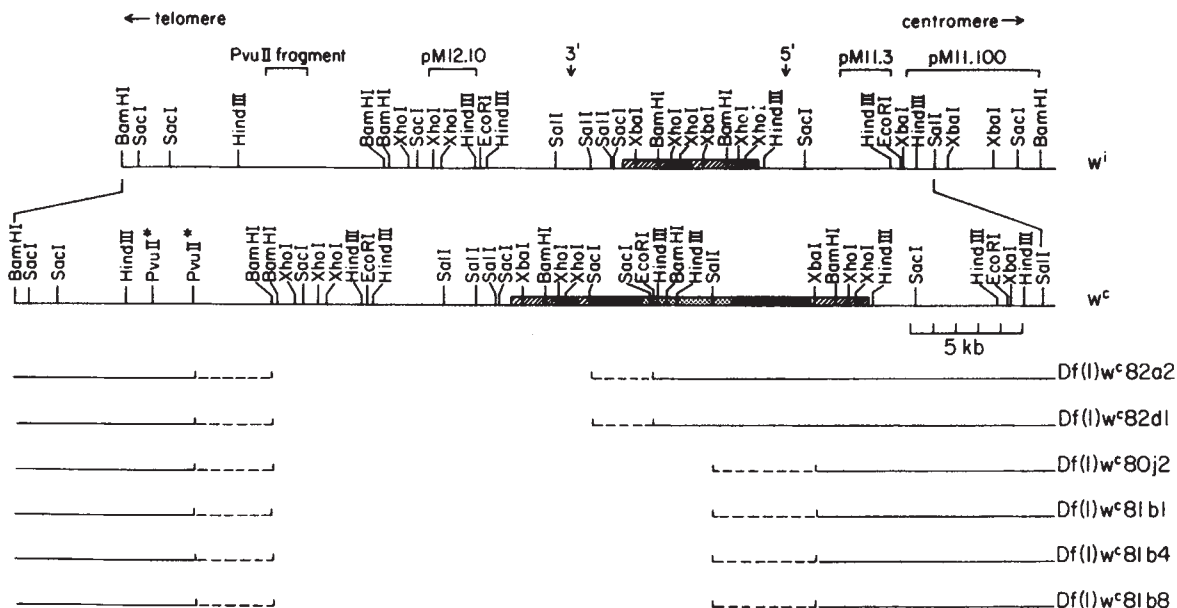
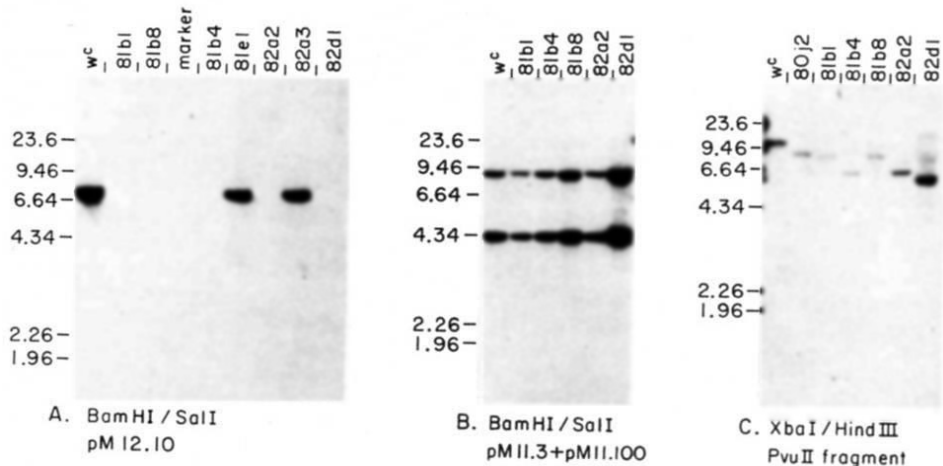


Fig. 1 Restriction maps of the *white* locus region from w^i , w^c and six deletion derivatives of w^c . Hatched bars on the w^i and w^c maps indicate regions duplicated in w^i . The solid bars represent the terminal repeats in the w^c insertion; the stippled bar represents the w^c middle, a sequence within the insertion which is not homologous to the inverted repeats. Maps of the deletion chromosomes are shown below the w^c map. Conserved regions of the deletion chromosomes are indicated by the solid lines drawn with respect to the w^c map; broken lines indicate uncertainty in the break points and deleted sequences are indicated by the absence of lines. Maps of the deletions were determined by DNA blots of digests of genomic DNA carried out as described previously⁹. The *SacI* site marking the left terminus of the w^c insertion is deleted in all six deletion chromosomes. The 5' and 3' limits of the wild-type *white* locus transcript are indicated by the arrows positioned above the w^i map²⁰. The two *PvuII* sites on the w^c map are shown with asterisks to indicate that *PvuII* sites on the w^c map have not been mapped throughout the entire DNA segment shown. The brackets above the w^i map show the location of hybridization probes used in experiments described. The *PvuII* fragment probe was isolated from a w^c background; all other probes are from a wild-type background¹². Derivatives of w^c were all isolated independently as spontaneous mutations from either a homozygous $y^2 w^c spl$ stock or from $y^2 w^c spl$ males carried with C(1)DX, $y w f$ females. Derivatives were maintained as hemizygotes or homozygotes on standard *Drosophila* medium⁹. All w^c derivatives are named with standard *Drosophila* nomenclature.

Fig. 2 Examples of blot hybridization of deletion chromosomes probed with sequences from wild-type *white* locus region. **a**, Deletion of sequences to the left of the element. Alleles w^{81el} and w^{82a3} are not deletions and are discussed below. **b**, Conservation of sequences to the right of the element in several deletion chromosomes. **c**, Size heterogeneity in deletion remnants. A continuation of these analyses was used to construct maps in Fig. 1. Genomic DNA was isolated from adult flies, digested as indicated under the blots, run on 0.6% agarose gels and transferred to nitrocellulose⁹. The limits of all four hybridization probes are shown with respect to w^i chromosome in Fig. 1. Size standards in kilobases are shown next to each figure, and are derived from *HindIII* digests of λ DNA run on the same gel.



A w^i triplication

Four of the 10 new *white* alleles we have analysed contain small rearrangements of sequences within or adjacent to the w^c insertion. For two alleles, hybridization of DNA blots with probes derived from the wild-type *white* locus indicated that *white* locus sequences had not been deleted. However, the presence of additional fragments with the size expected for the novel junction produced by the w^i duplication indicated that these chromosomes now contained a third copy of the sequence which is duplicated in w^i and w^c (Fig. 4). For one of these chromosomes, w^{81el} , restriction sites within the w^c insertion are conserved, indicating that the w^c insertion is intact and its position unaltered with respect to flanking DNA. However, the insertion is now flanked on the left side by two copies of the sequence duplicated in w^i ; a single copy remains on the right side of the insertion.

The simplest mechanism for generating this rearrangement is by unequal crossing-over. Because the original w^c insertion

occurred slightly to the right of centre in the w^i duplication, there is a small duplication on the left side of the w^c insertion. This is indicated on the w^c chromosome by the positions of the *PvuI* and *PvuII* sites in Fig. 4. Unequal crossing-over within this small duplication would result in the structure observed for w^{81el} . We isolated this mutation as a cluster of exceptional progeny from a stock in which only males carried the w^c chromosome. Thus, unequal crossing-over would have occurred as a pre-meiotic sister chromatid exchange event.

The second w^i triplication chromosome, w^{82a3} , differs from w^{81el} by an additional rearrangement event. In this chromosome, 1 kb of DNA has been deleted from the w^c insertion, with most or all of this deletion occurring within the left inverted repeat. The *SacI* site which precisely marks the left terminus of the insertion is conserved, as are sites within the insertion, but restriction fragments which span the left inverted repeat are 1 kb shorter than in w^c (Fig. 4). Thus, two events were

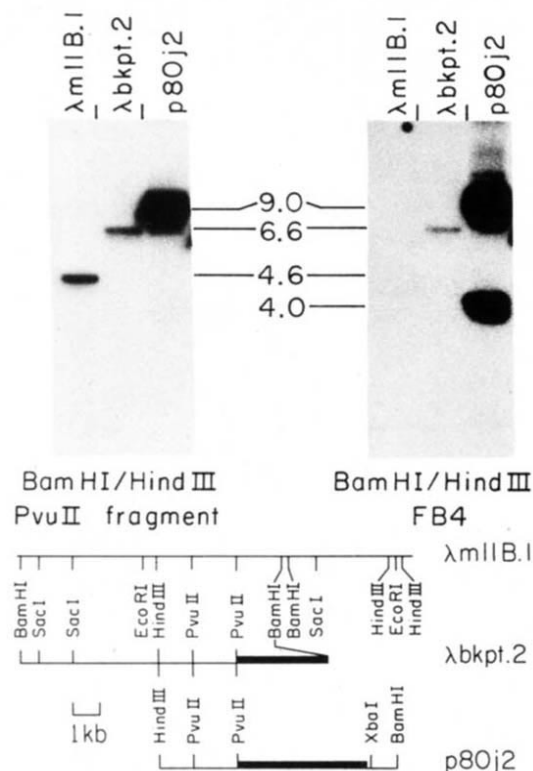


Fig. 3 Blots and restriction maps of isolated deletion break-point sequences in wild type, w^c and $Df(1)w^{80j2}$. Phage λ m11B.1 was previously isolated and contains sequences from a wild-type Canton S background¹². λ bkpt.2 contains a *Bam*HI fragment isolated in the λ phage vector Charon 28 (ref. 21) using methods described elsewhere⁹. This sequence was isolated from a w^i revertant of w^c , a chromosome which differs from the w^c chromosome only by the precise excision of the w^c insertion¹⁰. p80j2 contains the *Bam*HI–*Hind*III fragment from the $Df(1)w^{80j2}$ deletion break point. This genomic fragment was isolated in the plasmid vector pBR322 (ref. 22), since we have found that this stabilizes large FB DNA insertions⁹. The *Drosophila* DNA insertion in this plasmid comigrates with the same fragment in genomic DNA on agarose gels (data not shown). The left-hand blot is a *Bam*HI–*Hind*III digest of all three cloned DNA sequences hybridized with the *Pvu*II fragment directly adjacent to the deletion break point (see Fig. 1 for position of this probe). All three sequences show hybridization, and the *Bam*HI–*Hind*III fragments in the Canton S (λ m11B.1) and w^c (λ bkpt.2) backgrounds differ by 2.0 kb. A second blot derived from the same gel and hybridized with FB4, a plasmid containing an FB element⁴, shows hybridization only to λ bkpt.2 and p80j2; homology to FB DNA is absent in the phage from the wild-type background. The FB4 probe also hybridizes to the pBR322 vector sequences in p80j2. Hybridization to plasmid sequences is stronger because more DNA was loaded on the gel. The limits of homology to FB DNA were determined by a continuation of this analysis, and are indicated by the solid bars on the restriction maps.

required to produce this new rearrangement. We cannot determine if these events occurred independently.

Despite the structural similarity of w^{81e1} and w^{82a3} , only w^{81e1} is apparently mutable. We have not detected any mutations of w^{82a3} in several thousand flies, although we have recovered w^c , w^i and wild-type derivatives from a similar number of w^{81e1} flies. This suggests that the deletion within the inverted repeat of w^{82a3} is stabilizing the insertion.

Origin of a w allele

A third rearrangement chromosome, w^{80k1} , arose by the partial deletion of sequences within the w^c insertion and by the deletion of one copy of the w^i duplication. Restriction enzyme sites within the insertion and in the left copy of the w^i duplication are absent, leaving a structure that is now equivalent to a 3.5-kb insertion in the *white* locus (Fig. 5).

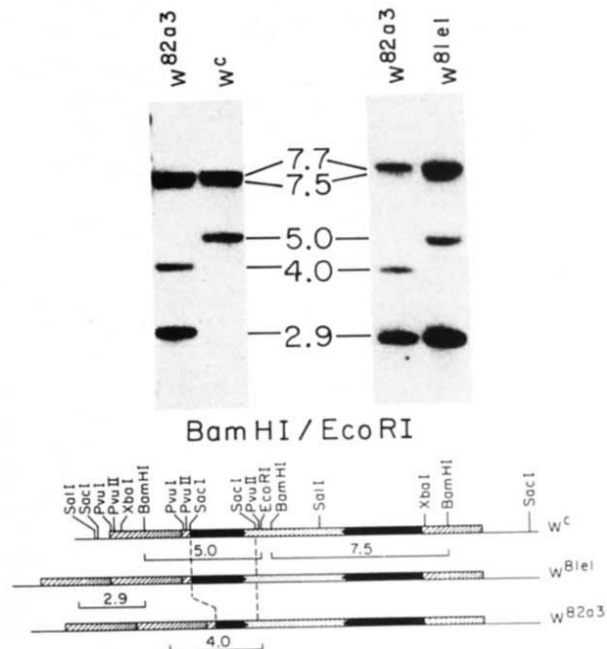


Fig. 4 Blots and restriction maps of w^c , w^{81e1} and w^{82a3} . The upper panels show blots of genomic DNA digested with *Bam*HI plus *Eco*RI. The hybridization probe is pM12.5, a subclone derived from the wild-type *white* locus which extends from the first *Sal*I site on the left of the w^i duplication to the *Hind*III site directly to the right of the w^i duplication¹². Restriction maps shown are derived from blots of genomic DNA. The hatched bars represent the sequences which are duplicated in w^i . The solid bars represent the inverted repeats within the w^c insertion; stippled bars indicate the w^c middle. Brackets below the maps show the positions of restriction fragments in the blot above. The 7.7- and 7.5-kb fragments are derived from sequences flanking the insertion, and from the right inverted repeat of the insertion, respectively. The 5.0-kb fragment spans the left inverted repeat in the w^c insertion. A deletion within this fragment produces the 4.0-kb fragment seen for w^{82a3} . The 2.9-kb fragment is derived from the extra copy of the w^i duplication in w^{81e1} and w^{82a3} . These and additional digests, blots and hybridizations were carried out as described previously⁹ to construct the restriction maps shown.

The location of a *Sac*I site at or near the break point in w^{80k1} suggested two possibilities for the origin of this allele. The *Sac*I site could have been created at the break point of a simple deletion with end points in FB DNA and *white* locus DNA, distinguishing this deletion from those previously described. Alternatively, this derivative could have arisen in two steps. The deletion of an internal portion of the w^c insertion would leave intact the *Sac*I site marking the left insertion terminus. An additional recombination event occurring within the small duplication present on the left side of the w^c insertion (represented by the *Pvu*II sites in Fig. 5) would eliminate sequences from one copy of the w^i duplication, and restore the sequence flanking the altered insertion to wild type.

We isolated a fragment containing the deletion break point of w^{80k1} in a recombinant plasmid, and determined the nucleotide sequence of portions of this fragment to distinguish between the two possibilities (Fig. 5). We first determined the DNA sequence of a fragment spanning the left edge of the insertion remnant which contains both *white* locus and FB DNA. The left terminus of the residual w^{80k1} insertion is identical to that of the original w^c insertion; the *Sac*I site which precisely marks the left terminus of the insertion is conserved, as are the first 31 nucleotides within the insertion (see ref. 11 for sequence). The *white* locus sequences which flank the insertion are also identical to those in the original w^c chromosome, indicating that this allele could not have arisen by a simple deletion event beginning within the element and extending leftward into *white* DNA. We also determined the sequence of the right terminus of the w^{80k1} insertion and flanking DNA, and found that it was

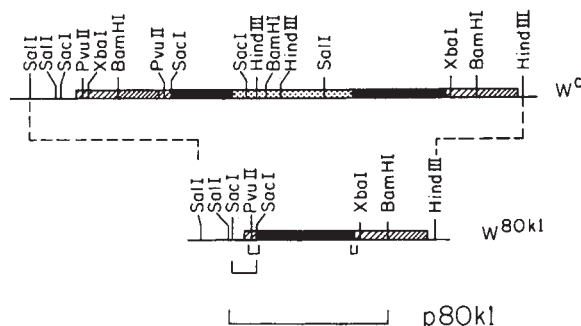


Fig. 5 Restriction maps of w^c and w^{80k1} . Hatched bars indicate the sequences duplicated in w^i . The solid bars show the location of the inverted repeats within the w^c insertion; the stippled bar represents the w^c middle. Restriction maps are derived from digests of genomic DNA and of the plasmid p80k1. p80k1 contains the $SacI$ - $BamHI$ fragment which spans the rearrangement in w^{80k1} , indicated by the large bracket below the w^{80k1} map. p80k1 was isolated directly from a plasmid library made from a $BamHI$ - $SacI$ digest of w^{80k1} DNA and the vector pBR322 (ref. 22) using methods described previously⁹. The restriction maps of this cloned sequence and the genomic sequence are co-linear. The solid bar on the w^{80k1} map shows the limits of hybridization of p80k1 to FB DNA. We determined the DNA sequence of the 800-bp $SacI$ fragment to the left of the insertion in w^{80k1} , and of the two $Sau3A$ - $TaqI$ fragments which overlap the edges of the insertion in w^{80k1} . These fragments, indicated by the small brackets under the w^{80k1} map, were isolated from the plasmid sequence in the M13 phage vector mp9, mp10 or mp11 (ref. 23), and the DNA sequence was determined by the dideoxynucleotide chain termination method²³⁻²⁵ as described previously¹⁰. The sequence of the $SacI$ fragment was determined on only one strand, and compared with the previously determined sequence from the wild-type locus (K. O'Hare, C. Murphy and G.M.R., in preparation).

conserved. Thus, a deletion has occurred within the element in w^{80k1} , reducing the size of the insertion from 10 kb to 3.5 kb.

Comparison of the DNA sequence of the 800-base pair (bp) $SacI$ fragment flanking the w^{80k1} insertion with the previously determined sequence of the same region in the wild-type *white* locus (K. O'Hare, C. Murphy and G.M.R., in preparation) revealed no differences between these two sequences (data not shown), indicating that deletion of one copy of the w^i duplication must have occurred by a homologous recombination event. Thus, w^{80k1} arose by two events, a deletion of sequences within the insertion and a recombination event in the DNA flanking the insertion. We cannot determine if these events occurred independently.

A stable white phenotype

The fourth rearrangement chromosome, w^{80j1} , apparently arose by a rearrangement of sequences within the w^c insertion. Sequences adjacent to the insertion are unaltered, as indicated by hybridization of genomic DNA blots to *white* locus probes. Restriction fragments having one end point within the *white* locus and one within the insertion are also conserved in size, indicating that the cluster of restriction sites within the middle of the w^c insertion is present, and that its position is unaltered with respect to flanking DNA. However, fragments which span the w^c insertion are ~4 kb larger for w^{80j1} than for the w^c allele (Fig. 6). These data can be simply interpreted by the assumption that w^{80j1} contains a tandem duplication of the w^c middle within the insertion.

Although the structure of w^{80j1} might suggest that it could revert to both w^i and wild-type phenotypes, we have never recovered any revertants from this strain. If a 'transposase' function were required for w^{80j1} mutability, one might expect to provide this function by adding a mutable copy of w^c to the w^{80j1} strain. We constructed a strain which had w^{80j1} on the X chromosome and a transposition of w^c on the third chromosome. This w^c transposition only mutates from crimson to white or light crimson phenotypes. We screened 9,376 flies and recovered

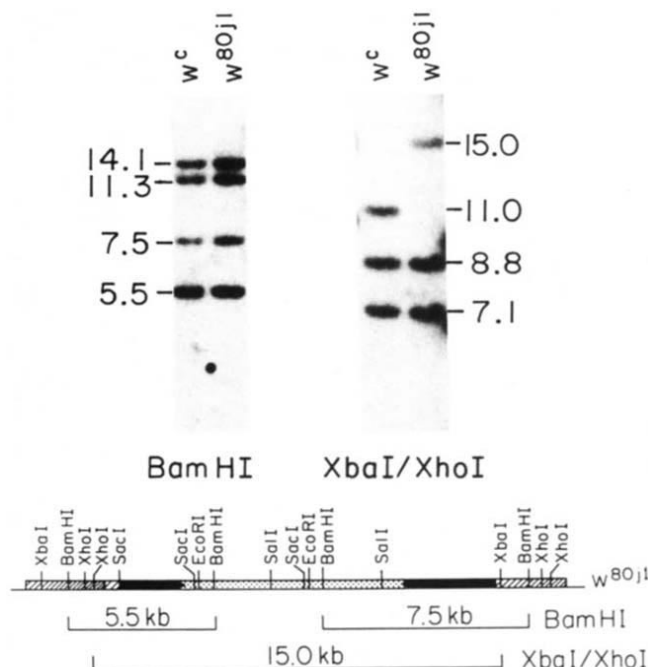


Fig. 6 Blots and restriction map of w^{80j1} . The restriction map of w^{80j1} is derived from blots of genomic DNA. Hatched bars illustrate the sequences which are duplicated in w^i . The solid bars represent the terminal repeats within the w^c insertion, and the stippled bar represents sequences derived from the w^c middle. The left panel of the blot shows a $BamHI$ digest of w^c and w^{80j1} DNA run on a 0.6% agarose gel, transferred to nitrocellulose and hybridized with pM12.5 (see Fig. 4 for the limits of this probe). The 14.1- and 11.3-kb fragments are derived from sequences flanking the insertion and are conserved in w^{80j1} . The 7.5- and 5.5-kb fragments are derived from sequences which have one site in the flanking *white* DNA and a second within the w^c insertion. These fragments are also conserved. The right panel shows an $XbaI$ - $XhoI$ digest hybridized with the same probe. In this case a 0.3% agarose gel was transferred in order to size large fragments accurately. The 11.0-kb fragment which spans the w^c insertion is absent in w^{80j1} , and is replaced by a 15.0-kb fragment. The 8.8- and 7.1-kb fragments are derived from sequences flanking the insertion. Small fragments derived from within the w^i duplication are not shown on this blot. Sizes were determined by comparison with $HindIII$ digests and $PvuII$ digests of DNA run on the same gel. All mapping of this allele was done by positioning sites within the insertion with respect to sites outside the insertion. Sites within the insertion cannot be directly mapped by genomic DNA blots due to the repetitive nature of the insertion.

9 mutations to a white phenotype, and 2 mutations to a light crimson phenotype, indicating mutation of the transposed allele, but recovered no mutations to a wild-type phenotype, which would have indicated mutation of w^{80j1} . Thus, in conditions where the transposed w^c allele is mutable, the w^{80j1} allele is at least 10-fold more stable.

Discussion

We have examined the structure of 10 chromosomal rearrangements induced by the FB transposable element in *Drosophila*. Six of these rearrangements are deletions which apparently arose by recombination between closely linked FB elements; sequences between these elements are deleted, and variously sized remnants of FB DNA remain at the deletion junctions. The absence of any deletions terminating in adjacent *white* locus DNA indicates that recombination between FB elements is a preferred mechanism for deletion formation. Reversion of the *white* locus mutation w^{DZL} also occurs by recombination between closely linked FB elements¹³. However, recombination between elements may not be the only mechanism for deletion formation, because a deletion with an FB element at only one break point in the parental chromosome has recently been identified (R. Levis, personal communication).

The deletion of DNA segments located between closely linked transposable elements by recombination appears to be a common property of transposable elements, and has been observed for prokaryotic elements and for *Ty1* in yeast^{14,15}. Our observations indicate that this is also the case for a eukaryotic element with long inverted terminal repeats. In yeast, a 13-kb DNA segment located between two *Ty1* elements has been observed to transpose to new genomic locations¹⁶. Transpositions of w^c have also been isolated¹⁷, and we have preliminary evidence that in one case, a w^c transposition was mediated by the mobilization of sequences between the two FB elements which interact in deletion formation.

In 9 of the 10 rearrangements examined, new FB element structures have been generated. For the six deletion chromosomes, the size and structure of the residual element apparently depend on the position of the recombination event within the elements. Thus, recombination between elements provides a mechanism for generating much of the heterogeneity seen within the FB element family. Two of the six deletion chromosomes we examined have residual FB elements which have lost the w^c middle. Thus, an internal segment can be deleted from an FB element by recombination with another element. This may have a bearing on our previous observation that the w^c middle is present in only a few of the FB elements within the genome, and that this number varies between strains of *Drosophila*².

In three cases, element-internal rearrangements have occurred, indicating that rearrangements within a single element also have a significant role in the generation of FB element heterogeneity. All three internally rearranged insertions have now lost the property of mutability, indicating that there is functional as well as structural heterogeneity within the FB element family. A similar heterogeneity has been observed for the P element family of transposable elements in *Drosophila*, where many elements have undergone internal deletions¹⁸.

Received 19 September 1983; accepted 5 January 1984.

1. Shapiro, J. A. (ed.) *Mobile Genetic Elements* (Academic, New York, 1983).
2. Levis, R., Collins, M. & Rubin, G. M. *Cell* **30**, 551–565 (1982).
3. Paro, R., Goldberg, M. L. & Gehring, W. J. *EMBO J.* **2**, 853–860 (1983).
4. Potter, S. S., Truett, M. A., Phillips, M. & Maher, A. *Cell* **20**, 639–647 (1980).
5. Truett, M. A., Jones, R. S. & Potter, S. S. *Cell* **24**, 753–763 (1982).
6. Potter, S. S. *Nature* **297**, 201–204 (1982).
7. Green, M. M. *Genetics* **56**, 467–482 (1967).
8. Karess, R. E. & Rubin, G. M. *Cell* **30**, 63–69 (1982).
9. Collins, M. & Rubin, G. M. *Cell* **30**, 71–79 (1982).
10. Collins, M. & Rubin, G. M. *Nature* **303**, 259–260 (1983).
11. Judd, B. H., Shen, M. W. & Kaufman, T. C. *Genetics* **71**, 139–156 (1972).
12. Levis, R., Bingham, P. M. & Rubin, G. M. *Proc. natn. Acad. Sci. U.S.A.* **79**, 564–568 (1982).
13. Levis, R. & Rubin, G. M. *Cell* **30**, 543–550 (1982).

These defective P elements are apparently able to transpose only when complemented by complete elements¹⁹.

Homologous recombination seems to have an important role in generating the derivatives we examined. Recombination between elements generated deletions, while recombination between duplicated sequences adjacent to the w^c insertion occurred in three small rearrangements. In each of these small rearrangements, the recombination events occurred within the 400 bp flanking the w^c insertion on the left, suggesting a linkage between these events and the FB element terminus. If chromosome breaks occur at the ends of FB elements, the repair of such breaks by recombination would generate such novel structures. Recombination is clearly stimulated by the w^c insertion, since the w^c mutation reverts to wild type several orders of magnitude more frequently than does w^i (ref. 7). Both events require a recombination event between the two copies of the w^i duplication to restore gene structure^{8,9}.

The analysis of derivatives generated by the w^c mutation indicates that the type and frequency of transposable element-induced rearrangements depend on both the structure of the transposable element itself and the organization of nearby sequences. Rearrangements within the w^c FB element resulted in a loss of genetic instability, indicating that not all FB elements are capable of generating rearrangements at the frequency observed for the w^c insertion. Furthermore, the presence of a closely linked FB element resulted in the frequent deletion of sequences adjacent to the w^c insertion, indicating that the distribution of elements will affect the type of rearrangement induced. The presence of a duplication flanking the w^c insertion also resulted in the generation of novel derivatives by unequal crossing-over events. These observations are particularly interesting when viewed in the context of genomic evolution.

We thank Christine Murphy for technical assistance. This work was supported by grants from the NIH and the American Cancer Society to G.M.R.; M.C. was a NIH postdoctoral fellow.

14. Kleckner, N. A. *Rev. Genet.* **15**, 341–404 (1981).
15. Liebman, S., Shalit, P. & Picologlou, S. *Cell* **26**, 401–409 (1981).
16. Stiles, J. I., Freidman, L. R., Helms, C., Consaul, S. & Sherman, F. *J. molec. Biol.* **148**, 331–346 (1981).
17. Green, M. M. *Genetics* **61**, 429–441 (1969).
18. O'Hare, K. & Rubin, G. M. *Cell* **34**, 25–35 (1983).
19. Engels, W. A. *Rev. Genet.* **17**, 315–344 (1983).
20. O'Hare, K., Levis, R. & Rubin, G. M. *Proc. natn. Acad. Sci. U.S.A.* **80**, 6917–6921 (1983).
21. Rimm, D. L., Horness, D., Kucera, J. & Blattner, F. R. *Gene* **12**, 301–309 (1983).
22. Bolivar, F. et al. *Gene* **2**, 95–113 (1977).
23. Messing, J. & Viera, J. *Gene* **19**, 269–276 (1982).
24. Sanger, F., Coulson, A. R., Barrel, B. G., Smith, A. J. H. & Roe, B. A. *J. molec. Biol.* **143**, 161–178 (1980).
25. Messing, J., Crea, R. & Seeburg, P. H. *Nucleic Acids Res.* **9**, 309–321 (1982).

Isotypic and allotypic variation of human class II histocompatibility antigen α -chain genes

Charles Auffray*, James W. Lillie, David Arnot, Dario Grossberger, Dietmar Kappes & Jack L. Strominger

Department of Biochemistry and Molecular Biology, Harvard University, 7 Divinity Avenue, Cambridge, Massachusetts 02138, USA

DNA sequences of four human class II histocompatibility antigen α chain DNA sequences (derived from cDNA and genomic clones representing DC1 α , DC4 α , DX α and SB α) are presented and compared to DR α and to mouse I-A α and I-E α sequences. These data suggest possible mechanisms for the generation of polymorphism and the evolution of the DR, DC and SB families.

THE major histocompatibility complex (MHC), located on chromosome 6 in man and chromosome 17 in the mouse, contains several groups of genes whose products are involved in the regulation of the immune response to self and non-self antigens (reviewed in refs 1–4). Among them, the class II anti-

gens (HLA-DR, DC and SB in man, I-A and I-E in the mouse) are thought to be the products of the immune regulation genes and serve as restriction elements in cell-cell interactions and antigen presentation to T lymphocytes. They are heterodimers consisting of a heavy (α) and a light (β) chain which both pierce the plasma membrane and are expressed primarily on B lymphocytes and macrophages. The extensive polymorphism of the class II antigens detected by serological, functional and

* Present address: CNRS Institut d'Embryologie, 94130 Nogent-sur-Marne, France.

DC-1 α (JY)	E D I V A D H V A - - G V N L Y Q F Y G P
DC-4 α λ DCH-9 (JY)	E D I V A D H V A S Y G V N L Y Q S Y G P S G Q Y S H E
DC-4 α pDCH1 (JY)	Q S Y G P S G Q Y S H E
DC-1 α HB20 (LB)	E D I V A D H V A S C G V N L Y Q F Y G P S G Q Y T H E
DC-1 α (LB)	E D I V A D - V A - - G V N L Y O F Y G P - G Q ^F - - E
	1 5 10 15 20 25

Fig. 1 Comparison of the amino-terminal sequences of DC α chains. The amino acid sequence derived from the cDNA clones HB20 and pDCH1 (ref. 16) and the genomic clone λ DCH-9 are compared to the partial amino acid sequence of the DC-1 α chains from the cell lines JY (DR4, w6) and LB (DRw6, w6) taken from ref. 20–21. Dashes indicate positions which were not assigned in the protein sequences.

structural studies appears to be mainly restricted to the β subunits. The availability of recombinant inbred strains of mice has allowed a subdivision of the mouse I region: the I-A subregion contains genes encoding A β , A α and E β , whereas the I-E subregion contains the gene encoding E α . In man, HLA-DR antigens constitute the predominant class II antigen and display similar function and tissue distribution as the products of the D region primarily defined by the mixed lymphocyte reaction. Multiple class II antigens have been defined through serological studies, such as the DC, MT, MB, LB and Te series. The strong linkage disequilibrium that they display with the HLA-DR specificities has not allowed a further dissection of the DR region. The best defined structurally of these additional class II antigens is the DC-1 antigen which was shown to be identical to MT1, MB1 and LB12 which are all found in linkage disequilibrium with DR 1, 2 and w6 (ref. 5, 6). Attempts to provide a more accurate typing in the mixed lymphocyte reaction by using primed T lymphocytes led to the discovery of another class II antigen called SB (secondary B cell antigen) which is encoded centromeric to HLA-D as defined by studies of recombinant families and deletion mutants^{7,8}. The SB antigen elicits a strong proliferative response in a secondary mixed lymphocyte reaction, provided that there is identity at the DR locus.

Recent advances in our knowledge of the multiplicity and structure of the class II antigens has been made through the advent of recombinant DNA technology. Thus a molecular map covering a substantial part of the murine I region has been obtained⁹. cDNA clones have been obtained for the α and β chains of the HLA-DR and DC antigens (refs 10–16, reviewed in ref. 4), and in this and a subsequent paper¹⁷ cDNA clones for the α and β chains of SB antigens will be described. We reported previously the isolation of a cDNA clone (pDCH1) corresponding to a DC α chain¹⁶. Further studies have shown that the I-A α chain is the murine homologue of the DC α chain¹⁸, both chains being polymorphic. Accordingly we have detected a restriction enzyme polymorphism of the DC α gene

which correlates well with the DR/MT disequilibrium groups¹⁸. The DC α probe also showed that the human haploid genome contains at least four related α chain genes which were localized on chromosome 6 using γ -ray induced deletion mutants¹⁹. These genes include those encoding the DR and DC α chains which have significant similarity in the second extracellular (α 2) and the transmembrane domains. One of the two additional α chain genes shows a high degree of sequence similarity with the DC α probe and was called DX α because it is not known whether it is an expressed gene and to which family of serologically defined antigens it might correspond. In order to investigate further the nature of the polymorphism of the DC α chain and the multiplicity of the class II antigen α chain genes, we have isolated cDNA clones and genomic clones using pDCH1 as a probe in cross-hybridization experiments. DNA sequence analysis of two alleles of the DC α chain provided evidence that they are complex alleles which differ at a large number of positions, mainly in the α 1 domain. Possible mechanisms for the generation of this allelic polymorphism are discussed. The structure of a class II antigen α chain different from but related to DR α and DC α has also been determined and shown to correspond to the SB α chain. The possible evolutionary relationship between three isotypic forms of class II antigen α chain (DR α , DC α and SB α) is discussed.

cDNA clones

The amino-terminal amino acid sequence of the DC-1 α chain from the LB cell line²⁰ (DRw6, w6) was the basis on which the cDNA clone pDCH1 (obtained from the JY cell line, DR4, w6) was identified as encoding a DC α chain, tentatively the DC-1 α chain. But as there was one difference between the two sequences at position 18, we had suggested it might correspond to an allele of the DC-1 α chain, namely the α chain of the DC-4 antigen which is found in linkage disequilibrium with DR4 and 7, whereas DC-1 is linked to DR1 2 and w6 (the JY and LB cell lines share DRw6). The comparison was based on a limited number of positions as the pDCH1 cDNA clone starts at codon 17 of the mature protein and the amino terminal sequence does not extend beyond codon 28 and there are a few gaps. Recently, a partial amino terminal sequence has been determined for the DC-1 α chain in the JY cell line as well, which is identical to that obtained from the LB cell line²¹.

Because the DC α and DX α genes are so closely related, it was important to determine to which gene the cDNA clone pDCH1 corresponds. The analysis of γ -ray induced deletion mutants has allowed us to distinguish between these two genes using the Southern blot technique¹⁹. Moreover, a restriction enzyme polymorphism of the DC α gene detected in HLA-DR homozygous typing cells makes it possible to distinguish three different alleles of the DC α gene which appear to correlate with the DR/MT linkage disequilibrium group^{16,22}. In contrast, the

Table 1 Pattern of differences between two DC α alleles and the DX α gene in the coding regions

	Nucleotides					Amino acids				
	A	B		C		A	B ¹		C	
SP	2/82 2.4%	3/82 3.7%	2	1	0	1/27 3.7%	1/27 3.7%	1	0	0
α 1	19/246 7.7%	30/246 12.2%	14	13	3	10/82 12.2%	18/82 21.3%	7	6	5
α 2	2/282 0.7%	8/282 2.8%	1	7	0	1/94 1.1%	3/94 3.2%	0	3	0
CP, TM, CY	4/155 2.6%	10/155 6.4%	5	5	0	3/51 5.9%	4/51 7.8%	2	2	0
Total	27/765 3.5%	51/765 6.7%	22	26	3	15/254 5.9%	26/254 10.2%	10	11	5

Number and percentage of positions (nucleotides or amino acids), domain by domain where: A, DC-4 α and DC-1 α are identical and DX α different; B, DC-4 α and DC-1 α are different (first column), number of these differences which are identities between DC-1 α and DX α (second column), DC-4 α and DX α (third column); C, all three sequences are different. The origin of the different sequences is indicated in legend to Fig. 4.

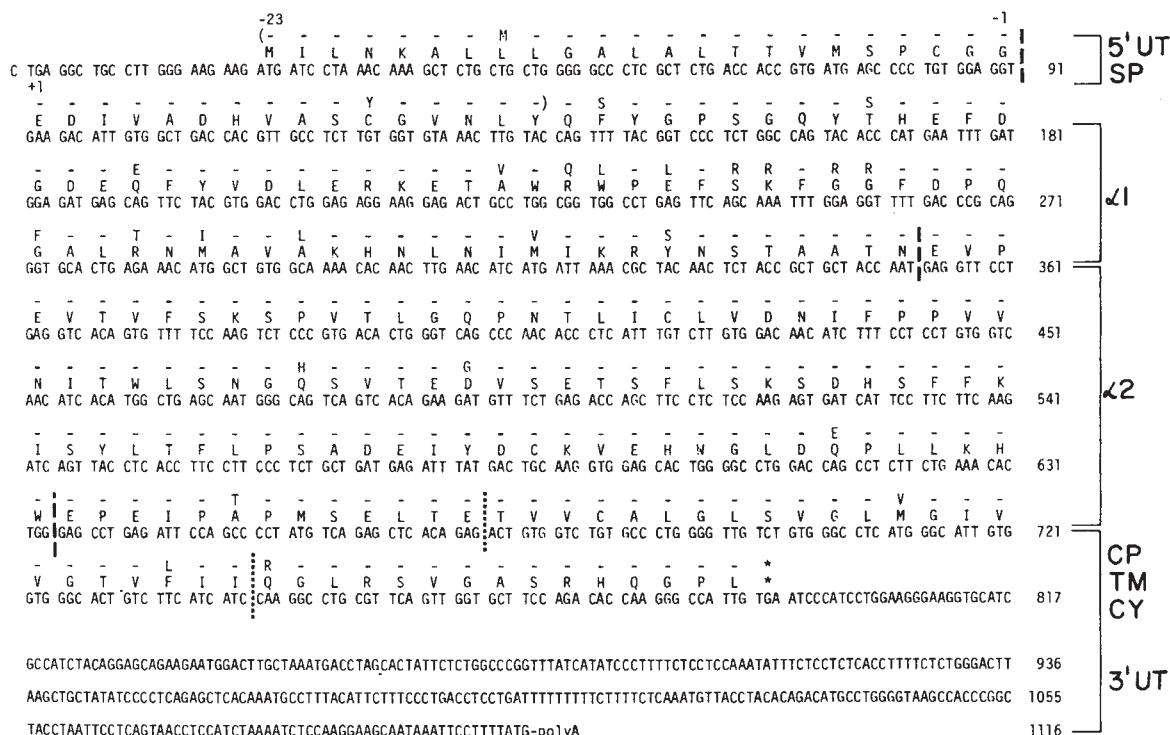


Fig. 2 Sequence of the HB20 cDNA clone. Total cellular RNA was extracted from the lymphoblastoid B cell line LB (DRw6, w6) using the LiCl-urea procedure³⁸. Poly(A) containing RNA obtained by oligo-dT cellulose chromatography was fractionated by sucrose gradient ultracentrifugation as described³⁸. Fractions containing the class II mRNAs were monitored by cell-free translation, immunoprecipitation and SDS-polyacrylamide gel electrophoresis. A cDNA library was constructed from the enriched fraction using conventional methods involving the insertion of double-stranded cDNA by dC-dG tailing in the *Pst*I site of the plasmid vector pBR322. Approximately 20,000 recombinant clones were screened using the pDCH1 insert as a probe in conditions described previously¹⁶. DNA sequence analysis was performed as follows. *Pst*I and *Sac*I fragments from the cDNA clone HB20 and LB18 (*DCα*) were isolated on a 2% agarose minigel by elution in wells cut in front of the bands visualized under UV light. The DNA was ethanol precipitated, then ligated to *Pst*I-cut pUC9 or *Sac*I-cut pUC12³⁹ for 2 h at 16 °C. The ligation mixture was used to transform *Escherichia coli* JM83 competent cells and white colonies were selected on plates containing 100 µg ml⁻¹ ampicillin in the presence of X-gal. DNA from the subclones was prepared from a 5 ml overnight culture by a boiling miniprep procedure⁴⁰. The DNA was cut with *Bam*HI or *Hind*III in the polylinker of the vector, labeled with the Klenow fragment of DNA polymerase I and [α -³²P]dNTPs, and recut with the second enzyme at the other end of the insert. These reactions were performed sequentially by heat-killing the restriction enzymes and adjusting the conditions by adding the appropriate buffer. The labelled fragment was isolated on a 2% agarose minigel as above, phenol extracted and ethanol precipitated. DNA sequencing reactions were performed according to Maxam and Gilbert⁴¹ and the products analysed on 80 cm 4% or 6% polyacrylamide-urea thin gels according to Sanger and Coulson⁴². Alternatively specific DNA fragments were labelled at their 3' end with terminal transferase and 3'dATP (cordycepin triphosphate) and isolated on 5% polyacrylamide gels by electroelution and DEAE-cellulose chromatography. The sequence of the HB20 clone is shown 5' to 3' together with its translation in amino acids using the single letter code. The sequence of the cDNA clone LB18 was identical to that of HB20. The top line is the amino acid sequence derived from pDCH1 for codons 17 to the carboxy terminus and from the λ DCH-9 genomic clone for the signal peptide and the first 16 codons of the α 1 domain¹⁶. Identical amino acids are indicated by dashes. The hexanucleotide AATAAA corresponding to a signal for polyadenylation is underlined. The delimitation of domains is indicated by vertical dotted lines according to the intron/exon organization of the *DRα* chain gene¹⁰. SP: signal peptide; α 1, α 2: first and second extracellular domains; CP, TM, CY: connecting peptide, transmembrane and cytoplasmic domains; 3'UT: 3' untranslated region of the mRNA.

DXα gene appears to be conserved by that method and is located on a 2.6 kb *Hind*III fragment in all haplotypes, whereas the *DC-4α* chain gene is located on a 5.5 kb *Hind*III fragment and a 2.4 kb *Pst*I fragment, and the *DC-1α* chain gene is located on a 9.5 kb *Hind*III fragment and a 11 kb *Pst*I fragment (ref. 16, 22 and unpublished observation). We constructed and screened a genomic library using JY-DNA in the bacteriophage λ vector L47 and isolated four strongly hybridizing genomic clones using pDCH1 as a probe. Two of them, λ DCH-1 and λ DCH-9 containing the 2.4 kb *Pst*I fragment corresponding to the *DC-4α* chain gene, were shown to be overlapping clones by restriction enzyme analysis. A partial sequence was derived from the genomic clone λ DCH-9, showing a complete identity with that of pDCH1, which therefore corresponds to the *DC-4α* chain. This analysis also allowed the determination of the complete amino terminal sequence of the *DC-4α* chain (Fig. 1), revealing one additional difference at position 11 between the *DC-4α* chain (tyrosine 11) and both JY and LB *DC-1α* chains (a hydrophilic amino acid at 11).

An mRNA fraction enriched for the α and β chain mRNAs

was used to construct a cDNA library from the LB cell line (DRw6, w6 and DC-1). In order to isolate sequences expressed in this cell line which cross hybridize with *DCα*, we screened the library with pDCH1 under non-stringent conditions (2 × SSC, 65 °C). A total of 25 positive clones were obtained and grouped in three categories according to their pattern of cross hybridization with probes specific for the *DRα*1 and *DRα*2 domains, the *DRα* 3' untranslated region (3'UT) and *DCα*. The major category contained 19 clones hybridizing strongly with *DRα*1, *DRα*2 and *DRα* 3'UT but weakly with *DCα*. Their restriction map was identical to that of authentic *DRα* clones^{10-12,15}. Most of these clones were full-length copies of the *DRα* mRNA, and the longest clone, pDRH7, contained a 1,300 base pair (bp) insert. Three clones hybridizing weakly with all probes are discussed later in this report and are shown to correspond to the *SBα* chain. The final category contains three clones which hybridized strongly with *DCα*, weakly with *DRα*2 and not at all with *DRα*1 or *DRα* 3'UT, that is they have all the characteristics of *DCα* related sequences¹⁶. Restriction enzyme analysis and DNA sequencing showed that they all

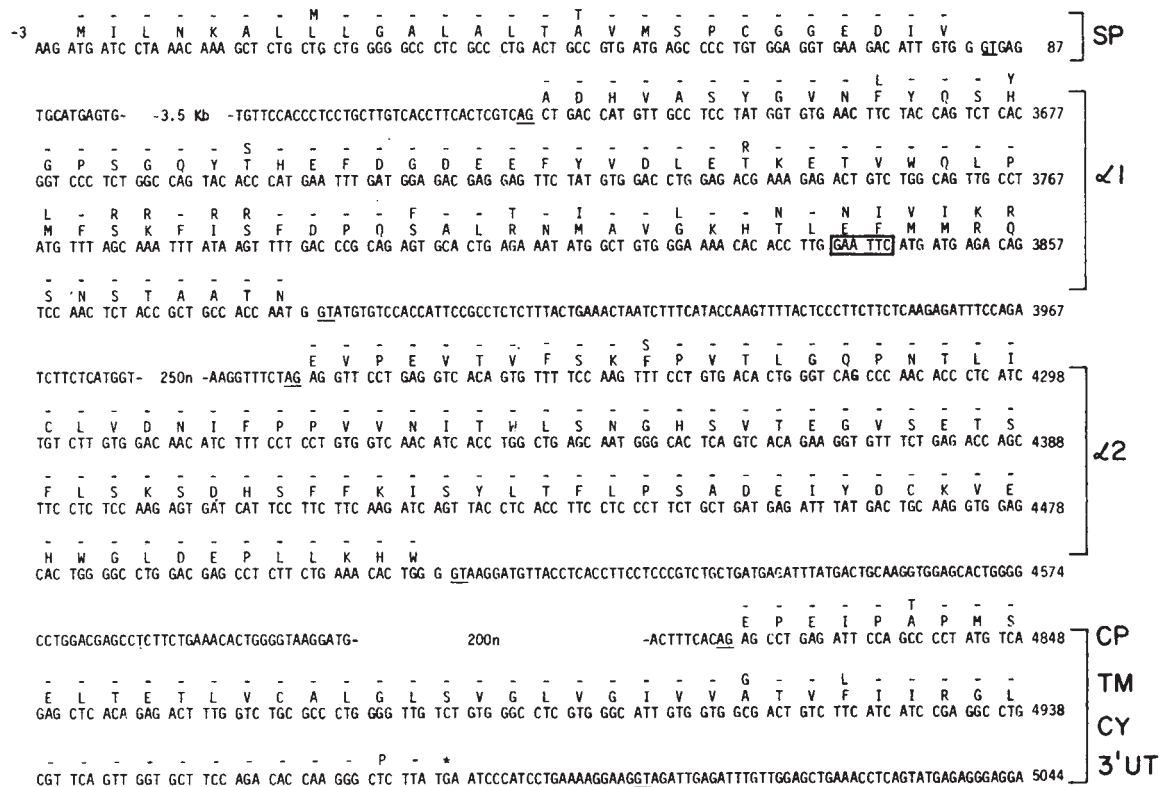


Fig. 3 Sequence of four exons of the $DX\alpha$ gene. High molecular weight DNA from the lymphoblastoid B cell line JY (DR4, w6) was partially digested with *Mbo*I and 20 kb fragments selected on a sucrose gradient, then ligated to the arms of λ L47 vector⁴³. A total of 400,000 clones were screened by plaque hybridization⁴⁴ with pDCH1 and four clones isolated as single plaques after two rounds of purification and rescreening. DNA fragments of the genomic clone λ DCH-10 ($DX\alpha$) and λ DCH-9 ($DC-4\alpha$) containing the $DC\alpha$ related sequences were detected by Southern blot hybridization, then subcloned in pBR322 or pUC9 (ref. 39). DNA sequence analysis was performed as described in legend to Fig. 2. The sequence of the four exons of the $DX\alpha$ gene is shown. Numbering starts at the ATG codon of the signal peptide and is indicated for the $\alpha 1/\alpha 2$ and transmembrane exons according to their approximate distance determined by restriction endonuclease analysis using the method of Smith and Birnstiel⁴⁵. The GT-AG dinucleotides which are part of the consensus sequences of splicing signals²⁸ are underlined. The translation of the different exons in amino acid is indicated using the single letter code. The upper line is the amino acid sequence of $DC-4\alpha$ deduced from pDCH1 (ref. 16), and λ DCH-9 for the signal peptide and the first 16 codons of the mature protein, with dashes indicating identical positions. The sequence corresponding to the *Eco*RI site found in the $DX\alpha$ 1 exon is boxed (see text).

correspond to the same sequence. The sequence of the longer cDNA clone containing a 1,200 bp insert (HB20) is shown in Fig. 2. The cloned sequence corresponds to a full-length copy of mRNA encoding a DC -like α chain, containing 22 bp of 5' untranslated region, a 23 amino acid hydrophobic signal sequence preceding the 232 amino acid open reading frame corresponding to the mature polypeptide chain followed by a 319 bp 3' untranslated region and the poly(A) tail located 11 nucleotides after the canonical AATAAA sequence²³. That the cDNA clone HB20 corresponds to an authentic $DC-1\alpha$ sequence is supported by the fact that the amino acid sequence derived from it is identical to the LB $DC-1\alpha$ sequence at all positions which were unambiguously assigned in the protein sequence (Fig. 1). In addition, the amino acids predicted from the nucleotide sequence at positions 7, 10, 22, 26, 27 would not have been detected with the methodology used. The two additional genomic clones hybridizing strongly with the pDCH1 probe, λ DCH-10 and λ DCH-11, were shown to be overlapping clones containing the 2.6 kb *Hind*III fragment specific to the $DX\alpha$ gene. An *Eco*RI site located in the $\alpha 1$ exon of the $DX\alpha$ gene was found to be conserved in several HLA-DR homozygous and heterozygous cell lines, including JY and LB (ref. 22 and unpublished observation). This *Eco*RI site is not found in HB20 which also shows extensive sequence differences with λ DCH-10, ruling out the possibility that HB20 could correspond to a transcript of the $DX\alpha$ gene (see below). We have therefore identified cDNA clones corresponding to the $DC-4$ (pDCH1) and $DC-1$ (HB20) α chain alleles.

Polymorphic $\alpha 1$ and Ig-like $\alpha 2$ domains

$DC-4\alpha$ and $DC-1\alpha$ alleles differ at 26 out of 254 amino acids and 51 out of 765 nucleotides (Fig. 2 and Table 1), a surprisingly high number for classical alleles which usually differ by one or a few substitutions. However, these differences are not distributed randomly. Most of them are found in the first extracellular domain ($\alpha 1$) whereas the second domain ($\alpha 2$) and the connecting peptide, transmembrane and cytoplasmic regions are much more conserved. Thus the extracellular portion of the $DC\alpha$ chain appears to consist of a 'polymorphic' $\alpha 1$ domain and a 'conserved' $\alpha 2$ domain. The $\alpha 2$ and $\beta 2$ domains of the α and β chains of the class II antigens show striking sequence similarity with several members of the immunoglobulin superfamily such as the immunoglobulin constant domains, β -2 microglobulin and the $\alpha 3$ domain of the class I antigens^{10,12,13,16,17,24-26}. This led to the hypothesis that they derive from a common ancestor. The diversity appears to be located in the amino terminal domain in both the α and β chains of class II antigens²⁷. Furthermore, a cluster of eight differences is found at position 45-56 of the $DC\alpha 1$ domain (Fig. 2), which has the characteristic of a 'diversity region' and could be part of an alloantigenic site. Although this pattern of variability is reminiscent of that of immunoglobulin heavy and light chain variable regions, the two gene families do not appear to have used a similar mechanism to generate it, since no somatic gene rearrangement or mutations have been detected in the class II genes.

Fig. 4 Comparison of the nucleotide sequence of the coding regions in two $DC\alpha$ alleles and $DX\alpha$. The sequences from which these data were compiled are shown in Figs 2 and 3 and in ref. 16. In addition, the signal sequence and the beginning of $\alpha 1$ of $DC4$ were derived from a genomic clone, $\lambda DCH-9$. The amino acid substitutions shown in Fig. 2 were derived from the following nucleotide sequence in $\lambda DCH-9$: ATG ATC CTA AAC AAA GCT CTG ATG CTG GGG GCC CTC GCC CTG ACC ACC GTG ATG AGC CCT TGT GGA GGT GAA GAC ATT GTG GCT GAC CAT GTT GCC TCT TAC GGT GTA AAC TTG TAC. ● indicates a position where $DX\alpha$ is identical to HB20, ○ a position where $DX\alpha$ is identical to pDCH1, ■ position unique to $DX\alpha$, and * position where all sequences differ. For the definition of the different domains of the α chain, see legend to Fig. 2.

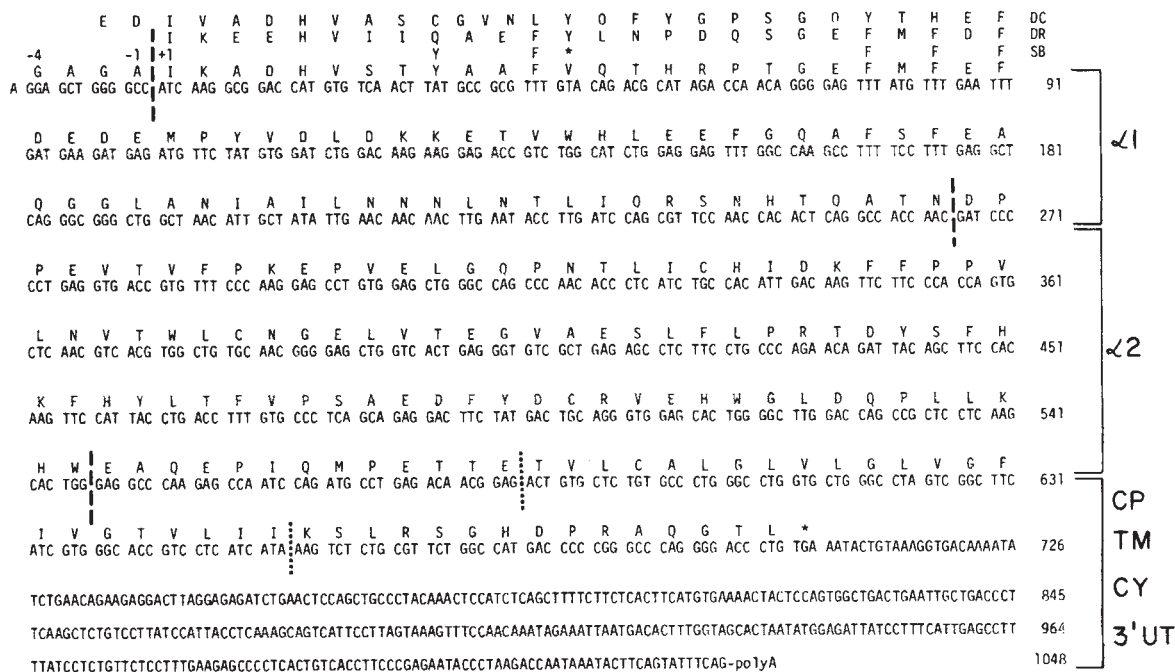


Fig. 5 Sequence of cDNA clones for the $SB\alpha$ chain. The cDNA clones LB14 and LB24 were obtained from the LB cDNA library and sequenced as described in the legend to Fig. 2. Both clones overlap and have identical sequences. The insert in LB14 corresponds to nucleotides 1–1,044 and the insert in LB24 to nucleotides 58–1,047 followed by a track of A residues. The translated sequence starts in the signal sequence at position –4 by analogy with $DR\alpha$ and $DC\alpha$ (refs. 10, 16). (The amino terminal sequences of the $DR\alpha$, $DC\alpha$ and $SB\alpha$ are taken from references 10, 16, 37 respectively.) The star in the $SB\alpha$ sequence indicates a position which is not Y or F^{37} , but is likely to be a valine (C. Katovitch-Hurley and J. D. Capra, personal communication).

Structure of $DX\alpha$ gene

In order to investigate further the nature of this complex allelism, we determined a partial DNA sequence of the $DX\alpha$ gene derived from the genomic clone $\lambda DCH-10$ (Fig. 3). The $DX\alpha$ gene contains at least four exons which potentially encode the signal peptide, $\alpha 1$, $\alpha 2$ and transmembrane domains of a class II antigen α chain. This structure is analogous to that described for the $DR\alpha$ chain gene^{10,26}. No abnormal feature found in this sequence would prevent the $DX\alpha$ gene from being expressed, that is, the splicing signals at the borders of the three introns follow the GT-AG rule and have typical consensus sequences²⁸, and there is no stop codon in the coding frame. Furthermore carbohydrate attachment sites are found at positions homologous to those found in the $DR\alpha$ and $DC\alpha$ chains^{10,16}, and the transmembrane region contains no charged residue. However, it is not excluded that unusual features are present in other parts of the gene which have not been sequenced.

The comparison of the amino acid and nucleotide sequences of $DX\alpha$, $DC-1\alpha$ and $DC-4\alpha$ shows that each sequence differs from either of the other two at approximately 25 amino acids and 50 nucleotides (Fig. 4 and Table 1). This is surprising as we would have expected the two $DC\alpha$ alleles to be more closely related to each other than to the $DX\alpha$ gene. Interestingly, at 48 of the 51 nucleotides at which the two $DC\alpha$ alleles differ, one of them is identical to the $DX\alpha$ sequence, that is there are

26 positions at which $DC-4\alpha$ is identical to $DX\alpha$, whereas at 22 others $DC-1\alpha$ is identical to $DX\alpha$. There are only three positions where all three sequences differ, all of which are located at the end of the $\alpha 1$ domain. In addition, there are 27 positions at which the $DC\alpha$ alleles are identical but $DX\alpha$ is unique, so that it would have been impossible to distinguish between allelic and non-allelic pairs on the basis of sequence comparisons.

There are several ways to interpret these observations. One possibility is that point mutations have been solely responsible for the observed differences so that at any position where one of the three sequences is unique, it is that particular sequence which has undergone mutation, the other two having conserved the nucleotide of the original common ancestor gene. In this case approximately 25 mutations in each gene could account for the some 50 differences between each sequence. Strong selective pressures following these point mutations in functionally important regions would then have resulted in their persistence in the population. The high frequency of these mutations observed at the $DC\alpha$ locus compared to the $DR\alpha$ locus^{10–12,15,26} might be related to the relative stability of the chromosomal regions containing these genes. Indeed in the mouse, the highly conserved $I-E\alpha$ gene is located telomeric to a hot spot for recombination, while the other genes, all polymorphic, are centromeric⁹. Another possibility, however, is that allelic forms of the $DC\alpha$ locus have not evolved independently of the $DX\alpha$ gene but that mutations accumulated in the $DX\alpha$

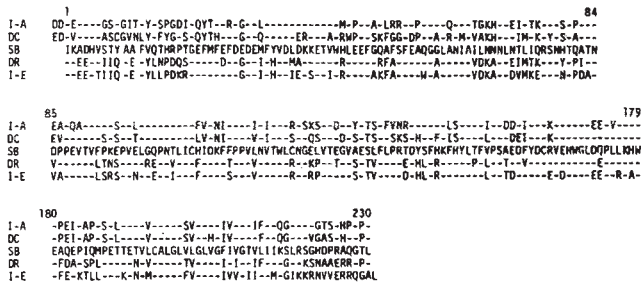


Fig. 6 Comparison of human and murine class II antigen α chains. The amino acid sequences are aligned domain by domain. A single gap has been introduced to maximize homology at position 14 in the SB α sequence and at position 13 in the DR α and Ea sequences. Residues which are identical at each position with the SB α sequence are indicated by dashes. The SB α sequence is that taken from Fig. 5. The DC α sequence is that of the HB20 cDNA clone (Fig. 2). (The DR α sequence is taken from ref. 10, the A α and Ea sequences are taken from refs 46 and 47, respectively.)

gene have also been transferred to the DC α gene. Such intergenic (and alternatively interallelic) genetic exchanges would have taken place by a copy repair mechanism analogous to gene conversion in yeast^{29,30}, which has been postulated to be operating on MHC class I genes as well as immunoglobulin genes³¹⁻³⁶. In both cases, the occurrence of point mutations would have established the original diversity. The occurrence of several blocks in the pattern of differences between the two DC α alleles and DX α (Fig. 4) may support the latter hypothesis and would substantiate the idea that gene conversion, which is primarily a mechanism leading to sequence homogenization, allows an amplification of the post-duplication divergence of alleles as in the mouse immunoglobulin γ_2a/γ_2b system³².

cDNA clones for SB α chain

Two cDNA clones (LB14 and LB24) corresponding to the third category of sequences cross-hybridizing with DC α (see above) were subjected to DNA sequence analysis. The sequence obtained is shown in Fig. 5 together with its translation in amino acids. Comparison with the DR α and DC α sequences reveals that this new sequence does not correspond to these α chain genes but rather to a third isotypic form of class II antigen α chain. The SB antigen has been defined functionally by its ability to stimulate primed T lymphocytes in a secondary mixed lymphocyte reaction, a function which can be blocked by the monoclonal antibody ILR1. This monoclonal antibody was subsequently used to show that the SB antigen has a structure typical of class II antigens, and a partial amino acid sequence has been reported for both the α and β chains of the SB3 antigen³⁷. Although there are only a limited number of positions available for comparison, they all match the amino acid sequence predicted from the LB14 cDNA clone (Fig. 5). In addition, the DNA fragments hybridizing strongly to the LB14 clone were lacking in γ -ray induced deletion mutants only when they had lost expression of the SB antigen (as detected with the ILR1 antibody) in addition to DC and DR (R. DeMars, personal communication). It should be kept in mind, however, that

evidence that the cDNA clones LB14 and LB24 encode the SB α chain relies primarily on the specificity of the ILR1 antibody. Preliminary analysis of a cosmid clone hybridizing strongly with both a SB β cDNA clone¹⁷ and LB14 has shown that it contains two SB α -related genes: SB α which hybridizes strongly with the probe and another gene provisionally named SX α , which hybridizes more weakly (unpublished observation, see also ref. 22). These genes do not correspond to any of the previously detected α chain genes¹⁹. The number of α chain genes is therefore at least five (DR α , two DC α -related and two SB α -related), and there are two additional genes or gene fragments designated DY α hybridizing weakly with a DC α probe¹⁹. Furthermore, it is possible that other α chain genes remain undetected with the DR α , DC α and SB α probes. This is in marked contrast with the situation in the mouse, where only two α chain genes have been found in the I region⁹.

Human α chains and murine homologues

The availability of the sequences of three isotypic forms of α chain allows the study of their evolutionary relationships. Such studies aim at understanding whether the SB antigen is a peculiarity of the human system which appeared late in evolution or whether it existed for a long time and was lost in the mouse, which has no known SB equivalent. On the basis of the analysis of the limited protein sequence data available and the cross-reactivity of the ILR1 monoclonal antibody with HLA-DR antigen, it has been argued that the SB α chain is an I-E or DR-like structure³⁷. The comparison of the complete sequences of the human DR, DC and SB and the murine I-A and I-E α chains clearly points out that DR α and DC α are the human equivalents of Ea and A α respectively. On the other hand, the SB α chain is very similar to all other α chains (54-61% matching) (Fig. 6). We have also made these comparisons domain by domain (Table 2). First, the SB α 1 domain is more closely related to DR α 1 than DC α 1 (56.5 vs 47%), but closer to A α 1 than Ea1 (57.6 versus 50.6%). Moreover, SB α 1 is as closely related to A α 1 as DC α 1 is (57.6 and 60%, respectively). It is therefore impossible to decide solely on the basis of this comparison if the human homologue of A α is DC α or SB α . The second relevant observation is that whereas DC α 2/A α 2 and DR α 2/Ea2 are very similar (72.3 and 80.8%), SB α 2 is as closely related to DC α 2 or DR α 2 (68.1 and 71.3%) as it is to A α 2 or Ea2 (63.8 and 67%). This comparison also does not allow a simple conclusion as to whether SB α is more like DC α or DR α . Finally, the pattern of homology observed in the third exon is different from that observed in the first two domains. Here SB α is clearly more like I-A α (56%) than like I-E α (39.2%), but it is as similar to DC α as to DR α (56.9 and 52.9%, respectively). A tentative explanation to account for these paradoxical observations would be that the three isotypic forms of α chain genes result from a succession of duplications of a common ancestor gene and that after each of these duplications, genetic exchanges similar to those we have invoked as a tentative explanation for the allelic polymorphism of the DC α chain have taken place, so that the isotypes appear as patchworks of each other. In any event, the gene duplication events creating the SB α chain gene ancestor appear to be old, leaving open the question of whether the mouse lost it or man gained it.

Table 2 Comparison of the amino acid sequences of human and murine class II antigen α chains

	α 1 (85 amino acids)				α 2 (94 amino acids)				CP/TM/CY (51 amino acids)				Total (230 amino acids)			
	DC	I-A	DR	I-E	DC	I-A	DR	I-E	DC	I-A	DR	I-E	DC	I-A	DR	I-E
SB	40	49	48	43	64	60	67	63	29	29	27	20	133	138	142	126
DC		51	41	37		68	59	60		46	30	19		165	140	116
I-A			45	38			55	56			29	21			129	115
DR				65				76				30				171

The number of identical amino acid positions in each domain is indicated for each combination of sequences. The number of amino acids in each domain is indicated. Excluded in this analysis are residues 1-2 in I-A α and DC α and residues 231 in I-E α since they have no equivalent in other sequences (see Fig. 6). The origin of the different sequences is indicated in legend to Fig. 6.

We thank Andrew Biro, Jeremy Boss, James Kaufman, Daniel Levy, Kiyotaka Okada, Magali Roux-Dosseto for useful discussion or providing reagents, Alice Furumoto-Dawson for manuscript editing, Roger Ollo and Francois Rougeon for communication of unpublished results and Ira Hierskowitz and Jonathan Seidman for critical reading of the manuscript. This

work was supported by grants from NIH (AM 30241) and the Kroc Foundation.

Note added in proof: The sequence derived from HB20 is also in perfect agreement with a recently reported partial amino acid sequence of DC-1 α ⁴⁸. Moreover, the sequence of a cDNA clone for DC-3 α from a DR 7 cell has now been reported⁴⁹.

Received 26 September; accepted 14 November 1983.

1. Strominger, J. L. et al. in *The Role of the Major Histocompatibility Complex in Immunobiology*, (ed. Dorf, M. E.) 115–172 (Garland, New York, 1981).
2. Benacerraf, B. *Science* **212**, 1229–1238 (1981).
3. Shackelford, D. A., Kaufman, J. F., Korman, A. J. & Strominger, J. L. *Immun. Rev.* **66**, 133–187 (1982).
4. Kaufman, J. F., Auffray, C., Korman, A. J., Shackelford, D. A. & Strominger, J. L. *Cell* **36**, 1–13 (1984).
5. Park, M. A., Terasaki, P. I. & Bernoco, D. in *Histocompatibility Testing 1980*, (ed. Terasaki, P. I.) 572–573 (UCLA, Los Angeles, 1980).
6. Shackelford, D. A., Mann, D. L., van Rood, J. J., Ferrara, G. B. & Strominger, J. L. *Proc. natn. Acad. Sci. U.S.A.* **78**, 4566–4570 (1981).
7. Shaw, S., Kavathas, P., Pollack, S., Charnot, D. & Mawas, C. *Nature* **293**, 745–747 (1981).
8. Kavathas, P., DeMars, R., Bach, F. H. & Shaw, S. *Nature* **293**, 747–749 (1981).
9. Steinmetz, M. et al. *Nature* **300**, 35–42 (1982).
10. Korman, A. J., Auffray, C., Schamboeck, A. & Strominger, J. L. *Proc. natn. Acad. Sci. U.S.A.* **79**, 6013–6017 (1982).
11. Lee, J. S., Trowsdale, J. & Bodmer, W. F. *Proc. natn. Acad. Sci. U.S.A.* **79**, 545–549 (1982).
12. Larhammar, D. et al. *Cell* **30**, 153–161 (1982).
13. Larhammar, D. et al. *Proc. natn. Acad. Sci. U.S.A.* **79**, 3687–3691 (1982).
14. Long, E. O. et al. *Proc. natn. Acad. Sci. U.S.A.* **79**, 7465–7469 (1982).
15. Wake, C. T. et al. *Proc. natn. Acad. Sci. U.S.A.* **79**, 6979–6983 (1982).
16. Auffray, C., Korman, A. J., Roux-Dosseto, M., Bono, R. & Strominger, J. L. *Proc. natn. Acad. Sci. U.S.A.* **79**, 6337–6341 (1982).
17. Roux-Dosseto, M. et al. *Proc. natn. Acad. Sci. U.S.A.* **80**, 6036–6040 (1983).
18. Auffray, C. et al. *EMBO J.* **2**, 121–124 (1983).
19. Auffray, C., Kuo, J., DeMars, R. & Strominger, J. L. *Nature* **304**, 174–177 (1983).
20. Bono, R. & Strominger, J. L. *Nature* **299**, 836–838 (1982).
21. Bono, M. R. & Strominger, J. L. *Immunogenetics* **18**, 453–459 (1983).
22. Trowsdale, J. et al. *Proc. natn. Acad. Sci. U.S.A.* **80**, 1972–1976 (1983).
23. Proudfoot, M. J. & Brownlee, G. G. *Nature* **263**, 211–214 (1976).

24. Orr, H. T., Lopez de Castro, J. A., Parham, P., Ploegh, H. L. & Strominger, J. L. *Proc. natn. Acad. Sci. U.S.A.* **76**, 4395–4399 (1979).
25. Long, E. O., Wake, C. T., Gorski, J. & Mach, B. *EMBO J.* **2**, 389–394 (1983).
26. Lee, J. S. et al. *Nature* **299**, 750–752 (1982).
27. Kaufman, J. F. & Strominger, J. L. *Nature* **297**, 694–697 (1982).
28. Breatnach, R. & Chambon, P. A. *Rev. Biochem.* **50**, 349–383 (1981).
29. Baltimore, D. *Cell* **24**, 592–594 (1981).
30. Radding, C. A. *Rev. Biochem.* **47**, 847–880 (1978).
31. Bentley, D. L. & Rabbitts, T. H. *Cell* **32**, 181–189 (1983).
32. Ollo, R. & Rougeon, F. *Cell* **32**, 515–523 (1983).
33. Evans, G. E., Margulies, D. H., Camerini-Otero, R. D., Ozato, K. & Seidman, J. G. *Proc. natn. Acad. Sci. U.S.A.* **79**, 1994–1998 (1982).
34. Pease, L. R., Schulze, D. H., Pfaffenbach, G. M. & Nathenson, S. G. *Proc. natn. Acad. Sci. U.S.A.* **80**, 242–246 (1983).
35. Lopez de Castro, J. A., Strominger, J. L., Strong, D. M. & Orr, H. T. *Proc. natn. Acad. Sci. U.S.A.* **79**, 3813–3817 (1982).
36. Weiss, E. H. et al. *Nature* **301**, 671–674 (1983).
37. Hurley, C. K., Shaw, S., Nadler, L., Schlossman, S. & Capra, J. D. *J. exp. Med.* **156**, 1557–1562 (1982).
38. Auffray, C. & Rougeon, F. *Eur. J. Biochem.* **107**, 303–314 (1980).
39. Vieira, J. & Messing, J. *Gene* **19**, 259–268 (1982).
40. Holmes, D. S. & Quigley, M. *Analyt. Biochem.* **114**, 193–197 (1981).
41. Maxam, A. & Gilbert, W. *Meth. Enzym.* **65**, 499–560 (1980).
42. Sanger, F. & Coulson, A. R. *FEBS Lett.* **87**, 107–116 (1978).
43. Loenen, W. A. & Brammer, W. J. *Gene* **10**, 249–259 (1980).
44. Benton, W. D. & Davis, R. W. *Science* **196**, 180–182 (1977).
45. Smith, H. O. & Birnstiel, M. L. *Nucleic Acids Res.* **3**, 2387–2398 (1976).
46. Benoist, C. O., Mathis, D. J., Kanter, M. R., Williams V. E. & McDevitt, H. O. *Proc. natn. Acad. Sci. U.S.A.* **80**, 534–538 (1983).
47. McNicholas, J., Steinmetz, M., Hunkapiller, T., Jones, P. & Hood, L. *Science* **218**, 1229–1232 (1982).
48. Götz, H. et al. *Hoppe Seyler's Z. physiol. Chem.* **364**, 749–755 (1983).
49. Chang, H.-C., Moriuchi, T. & Silver, J. *Nature* **305**, 813–815 (1983).

Alternative transcription and two modes of splicing result in two myosin light chains from one gene

Yo-ichi Nabeshima^{*,†}, Yoshiaki Fujii-Kuriyama[†], Masami Muramatsu[‡] & Kikuo Ogata^{*}

^{*} Department of Biochemistry, Niigata University School of Medicine, Niigata 951, Japan

[†] Department of Biochemistry, Cancer Institute, Japanese Foundation for Cancer Research, Kami-Ikebukuro, Toshima-ku, Tokyo 170, Japan

[‡] Department of Biochemistry, The University of Tokyo Faculty of Medicine, Hongo, Bunkyo-ku, Tokyo 113, Japan

Chicken skeletal muscle myosin alkali light chains are encoded by a single gene of size 18 kilobases (kb). This gene has two transcription initiation sites from which 17.5- and 8-kb precursor RNAs are transcribed. These RNAs are processed by different modes of splicing to form mRNAs encoding distinct light-chain (LC₁ and LC₃) proteins.

CHICK muscle cells have been used widely as a model system for the study of the control of gene expression during cell differentiation^{1–3}.

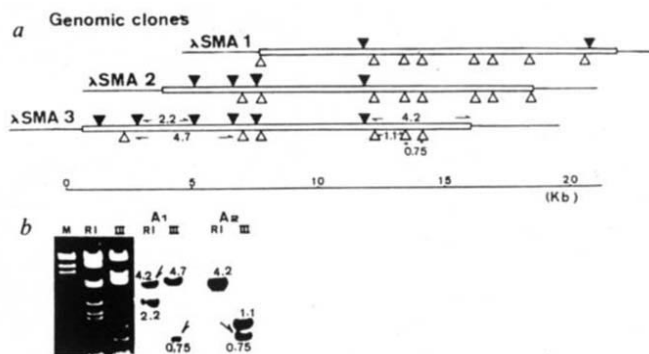
Myosin is a major protein component in the contractile apparatus in the cell, consisting of two large subunits of molecular weight 200,000 (200K) each and at least four smaller ones, that is, myosin alkali-released light chains and dithionitrobenzene (DTNB)-released light chains⁴. The amino acid sequences of these chick myosin light chains have been determined⁵. Skeletal muscle myosin is composed of two alkali light chains LC₁ (A₁) and LC₃ (A₂). Cardiac muscle has one alkali light chain which is homologous to skeletal muscle LC₁, while gizzard muscle has another alkali light chain which is homologous to skeletal muscle LC₃. As for the DTNB light chains (LC₂), skeletal and gizzard muscles have one DTNB light chain. Cardiac muscle, however, has two

homologous DTNB light chains, LC_{2A} and LC_{2B}. The sequence homology of these peptides suggests that genes of these alkali and DTNB light chains are related and have possibly evolved from one ancestral gene. Furthermore, during gene evolution, these genes must have acquired the machinery required for tissue-specific gene expression. The analysis of fine structure and regulatory mechanisms of this gene family should help to elucidate the control mechanisms of gene expression during cell differentiation. There have been several reports on the gene structure of myosin heavy and light chains^{6–10}.

To study light-chain gene expression, we have determined the nucleotide sequences of cloned cDNAs for skeletal muscle myosin alkali light chains LC₁ and LC₃ (ref. 11). The results indicated that not only the sequences of the coding region corresponding to the C-terminal 141 amino acid residues of LC₁ and LC₃ chains, but also those of the 3'-untranslated region of both mRNAs were completely identical. In contrast, the 5'-terminal 108 nucleotides of LC₁ were quite different from the 5'-terminal 78 nucleotides of LC₃.

§ Present address: Friedrich Miescher Institute, Basel, Switzerland.

Fig. 1 Restriction and hybridization analysis of cloned LC₁/LC₃ gene fragments. A chick gene library in Charon 4A bacteriophage prepared from random partial *Hae*III/*Alu*I digests of chick DNA was given by Dr J. D. Engel²⁹. Phage plaques were screened by the procedure of Benton and Davis³⁰ with the cloned chick skeletal muscle LC₁ and LC₃ cDNA inserts from plasmids pSMA₁-1 and pSMA₂-1 (ref. 11) as probes. Filter hybridization was performed overnight at 65 °C in 50 mM Tris-HCl buffer pH 7.5, containing 1 M NaCl, 10 mM EDTA, 0.1% sodium dodecyl sarcosinate, 0.2% polyvinyl pyrrolidone, 0.2% Ficoll and 0.2% bovine serum albumin. The filters were washed twice with 0.1×SSC (1×SSC = 0.15 M NaCl, 0.015 M sodium citrate) and 0.1% sodium dodecyl sarcosinate at 65 °C for 30 min. Phage DNA was prepared from lysates of *Escherichia coli* DP50/*supF* essentially as described elsewhere³¹. **a**, Cloned LC₁/LC₃ gene fragments (λSMA₁, λSMA₂ and λSMA₃) were mapped by single and double digestion of phage DNA. Phage DNAs (1 μg) were digested with *Eco*RI, *Hind*III, *Kpn*I, *Bgl*II, *Xho*I or *Bam*HI as recommended by the supplier. Restriction fragments were analysed by agarose gel electrophoresis followed by staining with ethidium bromide. Sizes of the fragments (given in kb) were determined with reference to the mobilities of *Hind*III-digested λDNA fragments. ▽, *Eco*RI; △, *Hind*III. Fragments detected by Southern hybridization in **b** are indicated by their length (in kb). **b**, λSMA₃ DNA was digested with *Eco*RI (RI) or *Hind*III (III) and electrophoresed on a 1% agarose gel. The restriction fragments were transferred to a nitrocellulose filter according to Southern¹². One set of filters was hybridized with 10⁷ c.p.m. ml⁻¹ of ³²P-labelled LC₁-unique (A₁) or LC₃-unique (A₂) segment at 65 °C and washed in stringent conditions. These segments were obtained from cloned cDNA (pSMA₁-1 and pSMA₂-1) by cutting with *Hae*III. 2.2-kb *Eco*RI and 4.7-kb *Hind*III fragments (A₁), and 4.2-kb *Eco*RI and 1.1-kb *Hind*III fragments (A₂) were demonstrated. Another set of filters was hybridized at 55 °C and washed with 5×SSC at 55 °C. In addition to the four bands described above, 4.2-kb *Eco*RI and 0.75-kb *Hind*III bands (A₁) and a 0.75-kb *Hind*III band (A₂) appeared (shown by arrows). Only the latter photograph is shown.



What is the relationship between the genes for the LC₁ and LC₃ proteins? There are two possibilities: (1) the LC₁ and LC₃ peptides are encoded by two distinct genes; (2) alternatively, while the 5'-terminal regions of the LC₁ and LC₃ mRNAs are encoded in two separate DNA segments, their 3'-terminal nucleotides are encoded by a common DNA segment on one gene. Complete identity of the 3'-side of the mRNAs strongly suggests the latter possibility. To distinguish between these possibilities, we isolated a genomic clone of myosin alkali light chain from a chicken gene library. Here we report the gene organization and a novel regulatory mechanism of synthesis of mature LC₁ and LC₃ mRNAs.

Cloning and characterization

Approximately 1×10⁶ plaques from the chicken gene library were screened with a ³²P-labelled LC₃ cDNA clone, pSMA₂-1 (ref. 11), as the probe. Ten clones which gave strong hybridiz-

ation signals were obtained, but these clones did not hybridize to the 5'-terminal region of an LC₁ cDNA clone, pSMA₁-1 (ref. 11). Therefore, two more genomic clones λSMA₂ and λSMA₃ (Fig. 1a), were obtained by using the LC₁-specific DNA fragment at the 5' terminus of pSMA₁-1 as probe.

To locate the sequences that specify the 5'-terminal portions of LC₁ and LC₃ mRNAs and the common region of both mRNAs, DNA fragments of these genomic clones were analysed by Southern blot hybridization¹² (see Fig. 1a, b). The region common to LC₁ and LC₃ mRNAs hybridized to a 4.5-kb DNA fragment localized at the right end of the λSMA₁ insert (data not shown). The 5'-terminal segment of LC₃ mRNA lay ~4.0 kilobases (kb) upstream from the common region of the two mRNAs, as shown by hybridization of a 1.1-kb *Hind*III fragment (Fig. 1b). In contrast, the 5'-terminal region of the LC₁ mRNA was found in a 2.2-kb *Eco*RI fragment (Fig. 1b) ~13 kb upstream from the region of the genome that the two mRNAs share (Fig. 1a). A 0.75-kb *Hind*III fragment hybridized to the

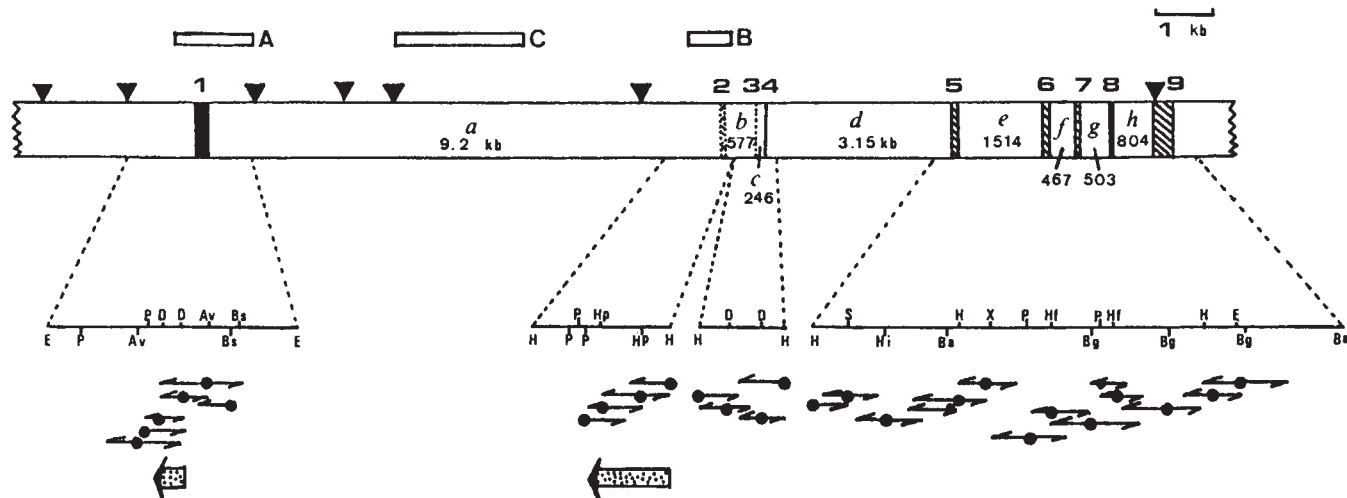


Fig. 2 Organization of the LC₁/LC₃ gene and sequencing strategy. Genomic DNA (21 kb) containing the LC₁/LC₃ gene is represented by a large horizontal bar, oriented in the 5' to 3' direction (left to right). The vertical bars numbered 1 to 9 on the restriction map indicate exons; *a-h* indicate intervening sequences, the lengths of which are also shown. A 2.2-kb *Eco*RI fragment (containing exon 1), *Hind*III fragments of 1.1 kb (exon 2), 0.75 kb (exons 3, 4) and 1.3 kb (exon 5) and a 3.5-kb *Bam*HI fragment (exons 6 to 9) were subcloned and mapped by single and double digestions of subcloned plasmid DNA. Ba, *Bam*HI; E, *Eco*RI; Bg, *Bgl*II; H, *Hind*III; Hf, *Hinf*II; P, *Pvu*II; X, *Xba*I; Hi, *Hinc*II; S, *Sau*96I; D, *Dde*I; Hp, *Hap*II; Bs, *Bst*EII; and Av, *Ava*I. Only those restriction enzyme sites that were used for DNA sequence determination are shown. Each arrow indicates the direction and length of the sequence determined from appropriate restriction sites. The exact locations of exons were determined by DNA sequence analysis. The lengths of these exons are: exon 1, 257 base pairs (bp); 2, 74 bp; 3, 25 bp; 4, 28 bp; 5, 144 bp; 6, 174 bp; 7, 78 bp; 8, 53 bp; 9, 317 bp. Stippled arrows show the DNA segments used for S₁ mapping of Fig. 4. The small open bars labelled A, B and C represent the DNA fragments used as hybridization probes for Southern blot analysis of skeletal muscle and liver DNAs (see Fig. 5). ▽, *Eco*RI site.

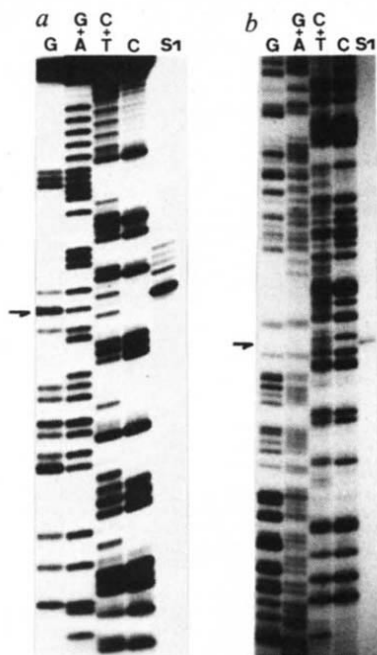


Fig. 4 S_1 nuclease protection mapping of the 5' termini of LC_1 and LC_3 mRNAs. *a*, 180-bp $DdeI/DdeI$ restriction fragment from a genomic clone, λ SM A_3 (see Fig. 2), extending from -172 to 8 of the LC_1 gene was prepared with ^{32}P -labelled 5' termini, then strand-separated. The single-stranded fragment (0.1 pmol) was hybridized with total poly(A) $^{+}$ RNA (40 μ g) of chick skeletal muscle in conditions described by Berk and Sharp¹³. After hybridization, the mixture was incubated with S_1 nuclease (100 U) and the digest analysed by 8% polyacrylamide-8 M urea gel electrophoresis. To provide chain length markers, the relevant DNA fragment was degraded by the Maxam-Gilbert procedure³² and electrophoresed in parallel, then autoradiographed for 24 h. *b*, A 1.1-kb $HindIII/HindIII$ restriction fragment of a genomic clone, λ SM A_3 , was prepared with ^{32}P -labelled 5' termini then cleaved with $PvuII$ (see Fig. 2). The resulting 0.62-kb $HindIII/PvuII$ fragment (0.1 pmol) was hybridized with the total poly(A) $^{+}$ RNA (250 μ g). The conditions for hybridization, S_1 digestion and gel electrophoresis were as described above. The labelled fragment which hybridized to the primary transcripts for LC_3 or LC_1 mRNA was protected from S_1 nuclease digestion. The content of the primary transcript was much smaller than that of mature LC_3 mRNA. As only the primary transcript from this region could protect this probe (note that the label is in the intron), the protected band was much weaker than in *a* where more abundant mRNA could protect the probe. The 0.62-kb protection band which hybridized to the primary transcript for LC_1 mRNA was also seen (data not shown).

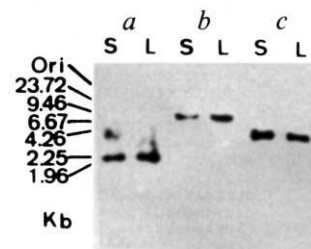
LC_1 and LC_3 5'-terminal regions in less stringent conditions (Fig. 1*b*). For further examination, these DNA fragments were subcloned in pBR322.

Coding and flanking regions

The DNA fragments containing coding sequences were identified by Southern blot hybridization as described above, then the nucleotide sequences of the exons and their flanking regions were determined as described in Fig. 2 legend.

The myosin alkali light-chain gene is separated into nine exons by eight intervening sequences (see Figs 2, 3). Exons 1 and 4 are specific for LC_1 , while exons 2 and 3 are specific for LC_3 . Exons 5 to 9 are shared by both LC_1 and LC_3 mRNAs. Exon 1 encodes the 5'-terminal noncoding sequence and N-terminal 42 amino acid residues of LC_1 . The second exon is present 9.2 kb downstream from exon 1. This exon 2 contains the 5'-terminal noncoding sequence of LC_3 mRNA and the initiation codon ATG. The third and fourth exons contain the coding sequences corresponding to 8 and 9 amino acid residues of LC_3 and LC_1 peptides, respectively. These two segments are connected precisely to the first nucleotide of the common region

Fig. 5 Analysis of LC_1/LC_3 gene in chick skeletal muscle and liver DNAs by Southern blot hybridization. High molecular weight chick skeletal muscle (S) and liver (L) genomic DNAs were digested with $EcoRI$. The products were electrophoresed on a horizontal 0.8% agarose slab gel, then transferred to a nitrocellulose filter by the procedure of Southern¹². The filters were hybridized to the nick-translated DNA fragment *a*, *b* or *c* (A, B and C in Fig. 2) at 65 °C for 16 h and washed with 0.1 $\times$ SSC. Specific activity of the probe was 0.5–1.5 $\times 10^8$ c.p.m. μ g $^{-1}$.



of LC_1 and LC_3 mRNAs present on exon 5 (Fig. 3*b*). This common region of LC_1 and LC_3 mRNAs is divided into five exons.

Location of 5' ends of LC_1 and LC_3 mRNAs

To localize the 5' end of LC_1 mRNA, a 5'-terminally labelled anticoding strand of the $DdeI/DdeI$ fragment (Fig. 2) was hybridized with the chick skeletal muscle poly(A) $^{+}$ RNA, and the reaction mixture digested with S_1 nuclease¹³. Figure 4*a* shows that one major protected band was detected in the gel together with some minor ones. The major protected fragment was 146 nucleotides long. The presence of a cap is known to hinder S_1 nuclease digestion of DNA probe by about four nucleotides as reported for β -globin¹⁴ and α -interferon¹⁵. We conclude, therefore, that the 5' terminus of the LC_1 mRNA lies at the G residue 131 nucleotides upstream from the initiation codon.

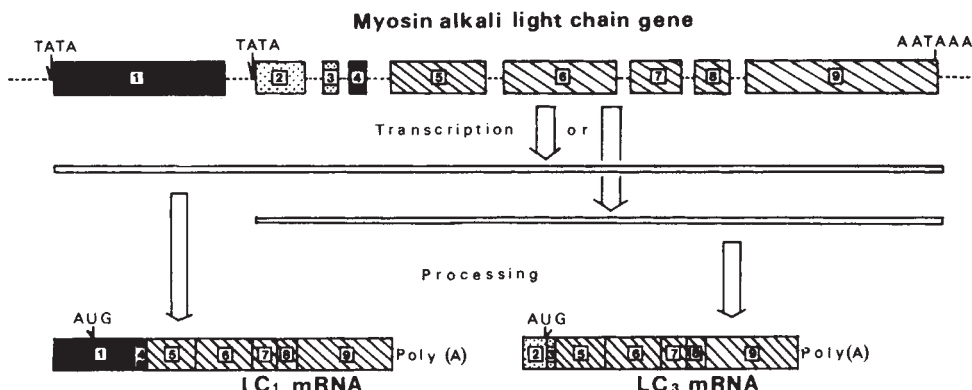
To determine the second initiation site, the $HindIII$ site 187 nucleotides downstream from the initiation codon of LC_3 mRNA and present in the intervening sequence, was labelled and cut with $PvuII$. The probe was hybridized with the total poly(A) $^{+}$ RNA, then a similar S_1 nuclease protection mapping was carried out to locate the 5' end of the LC_3 mRNA. Figure 4*b* shows that the major protected fragment was 269 nucleotides long. We conclude from these results that the 5' terminus of the LC_3 precursor RNA lies at the A residue 71 nucleotides upstream from the initiation codon.

Approximately 30 nucleotides upstream from the 5' terminus of LC_1 mRNA is a 'TATA box'¹⁶, TATATATA. A CCAAT sequence is also present ~70 nucleotides upstreams from the 5' terminus of LC_1 mRNA (Fig. 3). Another TATA box, ATAAATAA, is also found ~30 nucleotides upstream from the 5' terminus of LC_3 mRNA. These results support the idea that the transcription of the precursors of LC_1 and LC_3 mRNAs initiates at two different points. Short sequences of CTCTCA $_G$ (underlined in Fig. 3) occur between the 5' terminus and the translation initiation codon of both mRNAs.

Divergence between light-chain mRNAs

Figure 3*b* gives the sequence at the relevant splice junctions for the two alkali light-chain cDNAs and the corresponding regions of the genomic clone. Examination of the sequences at the 5' boundary of the first common exon (exon 5) and the 3' boundaries of LC_1 - and LC_3 -specific exons (exons 4 and 3 respectively) indicates that the divergence between LC_1 and LC_3 mRNAs occurs precisely at the splice junctions. The observation that a GT follows exons 3 and 4, and AG precedes exon 5 in the genomic clone, is in accordance with the GT-AG rule for a splice junction¹⁷ and is consistent with the prediction that each exon can be alternatively spliced to form the mature mRNAs. These results indicate that, whereas the 5'-terminal regions of LC_1 and LC_3 mRNAs are encoded in different DNA segments, their 3'-terminal sequences corresponding to the C-terminal 141 amino acid residues and the 3'-untranslated regions are encoded in a common DNA segment on one gene.

Fig. 6 Schematic representation of the LC_1/LC_3 gene and the mode of processing of LC_1 and LC_3 mRNA precursors. Boxes 1 to 9 show the exons and the dotted line represents the intervening sequences. Solid boxes (1, 4) and stippled boxes (2, 3) are unique for LC_1 and LC_3 mRNAs, respectively. Hatched boxes (5–9) are common to both mRNAs. TATA, TATA box. AATAAAA, putative polyadenylation signal. AUG, transcription initiation codon.



Structural rearrangement of LC_1/LC_3 gene

There are several possible mechanisms by which each mature mRNA can be produced. Because each unique sequence of mRNA is located 5' to the common sequence of LC_1 and LC_3 mRNAs in the gene, it is possible that the gene segment unique to LC_1 (exon 1) or LC_3 (exons 2+3) mRNA translocates to the upstream region of exon 4 or exon 5 by a gene recombination mechanism such as that demonstrated in immunoglobulin genes¹⁸. A second possibility is that the DNA segment of exon 1 or exons 2 and 3 binds to exon 4 or to exon 5 without deletion of any DNA segment by an unknown mechanism, then the precursor RNA is read through. As a third possibility, ~18- and 8-kb precursor RNAs may be transcribed, respectively, for LC_1 and LC_3 mRNAs and the latter two may be produced by different splicing pathways.

To test the first possibility, skeletal muscle and liver DNAs were examined for the organization of the LC_1 - and LC_3 -specific sequences. Figure 5 shows that both chick skeletal muscle and liver DNAs gave rise to the same single band of 2.2, 9 and 4.3 kb each by Southern blot hybridization of *EcoRI* digests, with the probes containing exon 1, exon 2 and the first intervening sequence, respectively (A, B and C in Fig. 2). The sizes of the resulting bands are the same for skeletal muscle DNA and nonproducer liver cell DNA and identical to those of the cloned genes (Figs 1, 5). These results strongly suggest that the skeletal muscle alkali light-chain gene occurs once per haploid genome and the skeletal muscle LC_1 and LC_3 mRNAs are transcribed from the same gene that we isolated from the gene library. The results also indicate the structural rearrangement does not occur in the region of the LC_1/LC_3 gene of the skeletal muscle cell DNA, and that the DNA segment corresponding to the first intervening sequence is not removed from the skeletal muscle gene during cell differentiation. The second possibility does not seem likely as in a preliminary experiment, we detected 17.5- and 8-kb RNA bands by Northern blot analysis¹⁹ of total poly(A)⁺ RNA of muscle cells, in agreement with the postulated sizes of the primary transcripts (data not shown).

Discussion

The present study has revealed a novel aspect of gene regulation in eukaryotic cells, that is, the production of two different polypeptides from one gene by differential transcription.

Some of the viral genes, including simian virus 40 (ref. 20) and adenovirus²¹, are known to produce more than one polypeptide from a single stretch of DNA. Apart from viral genes, a few cases are known where splicing occurs to produce different mRNAs. The splicing of alternative exons results in the expression of either the secreted or membrane type of immunoglobulin μ chain²² and δ chain²³, where a polyadenylation site intercedes between the exons involved in switching. The calcitonin gene is also reported to have alternative splice events, which results in an additional peptide, referred to as calcitonin gene-related peptide (CGRP)²⁴. More recently, the murine α crystalline gene was reported to produce two polypeptides by alternatively selecting an exon present in an intron²⁵. In the

amylase gene (*Amy1^A*), different transcription initiation sites of the same gene result in a different pattern of splicing, to produce different mRNAs; this, however, does not affect the structure of synthesized proteins²⁶.

This represents the first report of a cellular gene which produces two polypeptides from a single gene by differential initiation of transcription followed by a different mode of splicing (Fig. 6). The mechanism of the differential splicing is unknown. There are at least two possibilities. First, transcription initiation at different sites may determine the unique splice site(s), probably through the structure of the precursor molecule. Second, different splicing machinery (enzymes, cofactors, snRNAs) may work for different precursor molecules.

As both alternative pathways are operative in adult muscle cells, the simplest model would be the first possibility. When exon 1 is transcribed, the donor site of this exon finds the acceptor site of exon 4 as partner. When exon 2 is transcribed, on the other hand, the donor site of this exon finds the acceptor site of exon 3 which has been overlooked by exon 1. Then, exon 3 disregards the acceptor site of exon 4 but takes exon 5 as a partner. Whatever the mechanism for this, the presence of the alternative splicing pathways as depicted in Fig. 6 argues against any simple scanning model from either the 5' or 3' side, in agreement with Kühne *et al.*²⁷. Rather, it suggests a mechanism in which the secondary or tertiary structure of the precursor molecule has an important role in selecting the splice site. The fact that the presence of the upstream sequence affects tremendously the pattern of splicing indicates a major role for the upstream sequence (and probably its higher order structure) in the recognition of correct splice sites.

The regulation of the production of LC_1 and LC_3 chains is of special interest. In early embryonic muscle (10–14-day embryo), LC_1 is predominantly synthesized, with negligible amounts of LC_3 . The latter appears in the 14–15-day-old embryo and is present in adult muscle in an amount comparable to LC_1 (ref. 28). How this switch occurs is presently unknown. Whether only transcription initiation at specific sites determines all the subsequent events, including the mode of splicing, or whether other factors such as splicing machinery, stability of mRNA and translational control take part in the production of different polypeptides remains to be determined.

Received 20 September; accepted 7 December 1983.

- Keller, L. R. & Emerson, C. P. Jr *Proc. natn. Acad. Sci. U.S.A.* **77**, 1020–1024 (1980).
- Gauthier, G. F., Lowey, S., Benfield, P. A. & Hobbs, A. W. *J. Cell Biol.* **92**, 471–484 (1982).
- Masaki, T. & Yoshizaki, C. *J. Biochem.* **76**, 123–131 (1974).
- Weeds, A. G. & Pope, B. *J. molec. Biol.* **111**, 129–157 (1977).
- Matsuda, G., Maita, T., Kato, Y., Chen, J. I. & Umegane, T. *FEBS Lett.* **135**, 232–236 (1981).
- Mahdavi, V., Periasamy, M. & Nadal-Ginard, B. *Nature* **297**, 659–664 (1982).
- Kavinsky, C. J. *et al. J. biol. Chem.* **258**, 5196–5205 (1983).
- Robbins, J., Freger, G. A., Chisholm, D. & Gillam, T. J. *J. biol. Chem.* **257**, 549–556 (1982).
- Robert, B. *et al. Proc. natn. Acad. Sci. U.S.A.* **79**, 2437–2441 (1982).
- Nudel, U. *et al. Nucleic Acids Res.* **8**, 2133–2146 (1980).
- Nabeshima, Y., Fujii-Kuriyama, Y., Muramatsu, M. & Ogata, K. *Nucleic Acids Res.* **10**, 6099–6110 (1982).
- Southern, E. M. *J. molec. Biol.* **98**, 503–517 (1975).
- Berk, A. & Sharp, P. A. *Cell* **12**, 721–732 (1977).
- Weaver, R. F. & Weissmann, C. *Nucleic Acids Res.* **7**, 1175–1193 (1979).

15. Nagata, S., Matei, N. & Weissmann, C. *Nature* **287**, 401–408 (1980).
16. Banoist, C. & Chambon, P. *Nature* **290**, 304–310 (1981).
17. Sharp, P. A. *Cell* **23**, 643–646 (1981).
18. Sakano, H., Huppi, D., Heinrich, G. & Tonegawa, S. *Nature* **280**, 288–294 (1979).
19. Thomas, P. S. *Proc. natn. Acad. Sci. U.S.A.* **79**, 5201–5205 (1980).
20. Ziff, E. B. *Nature* **287**, 491–499 (1980).
21. Nevins, J. R. *Cell* **28**, 1–2 (1982).
22. Early, P. *et al. Cell* **20**, 313–319 (1980).
23. Maki, R., Roeder, W., Trauneker, A., Raschke, W. & Tonegawa, S. *Cell* **25**, 353–365 (1981).

24. Amara, S. G., Jonas, V., Rosenfeld, M. G., Ong, E. S. & Evans, R. M. *Nature* **298**, 240–244 (1982).
25. King, C. R. & Piatigorsky, J. *Cell* **32**, 707–712 (1983).
26. Young, R. A., Hagenbuchle, O. & Schibler, U. *Cell* **23**, 451–458 (1981).
27. Kühne, T., Wieringa, B., Reiser, J. & Weissmann, C. *EMBO J.* **2**, 727–733 (1983).
28. Roy, R. K. & Sarkar, S. *FEBS Lett.* **149**, 22–28 (1982).
29. Dodgson, J. B., Strommer, J. & Engel, J. D. *Cell* **17**, 879–887 (1979).
30. Benton, W. D. & Davis, R. W. *Science* **196**, 180–182 (1977).
31. Maniatis, T. *et al. Cell* **15**, 687–701 (1978).
32. Maxam, A. M. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **74**, 560–564 (1977).

LETTERS TO NATURE

Limits on surface magnetic fields of fast pulsars

B. Datta

Indian Institute of Astrophysics, Bangalore 560 034, India

The rapid rotation rate of millisecond pulsar¹ PSR1937+214 suggests that it may be close to the point of rotational instability. Rotational stability implies lower limits on the mass and moment of inertia, and an upper limit on the radius of the rotating neutron star. I show here that if fast pulsars are remnants of galactic bulge X-ray sources^{2–4}, the limits on the structure of rotating neutron stars, as implied by the requirement of rotational stability, can be used to set limits on the surface dipole magnetic field and the slow-down rate.

Neutron stars have essentially flat density profiles and are well approximated by Maclaurin spheroids. Above a certain rotational speed, there can be bifurcation from the axisymmetric Maclaurin sequence to non-axially symmetric Jacobi ellipsoids, at which point the star may become unstable. This (secular instability) limit corresponds to an angular speed Ω_s given by⁵

$$\Omega_s^2/2\pi G\langle\rho\rangle = 0.18 \quad (1)$$

where $\langle\rho\rangle$ is the density of the Maclaurin spheroid, and G is the gravitational constant. Assuming PSR1937+214 to be a uniformly rotating neutron star on the verge of secular instability, equation (1) implies (corresponding to its period, $P = 1.5577$ ms) the following minimum average density: $\langle\rho\rangle = 2.4 \times 10^{14}$ g cm⁻³. Once rotating neutron star configurations are known, this gives lower limits on the mass and moment of inertia ($M_{\min}$ and $I_{\min}$, respectively) and an upper limit on the radius ($R_{\max}$) of fast pulsars. For a representative choice of the equation of state (EOS) of neutron star matter. Datta and Ray have constructed rotating neutron star models⁶; these limits are summarized in Table 1. Also listed are $M_{\max}$, $I_{\max}$ and $R_{\min}$, obtained using the condition of hydrostatic equilibrium.

Galactic bulge X-ray sources invoke a neutron star accreting matter from a keplerian accretion disk fed by a companion. When the accreting region nearest to the neutron star is a keplerian disk, the neutron star will gain in angular speed until the keplerian angular velocity at the disk inner edge (the stellar magnetosphere's Alfvén radius) approximately equals that of the star. The equilibrium period when the steady state is obtained is^{7–9}

$$P \geq (6 \times 10^{-4}) B_8^{6/7} R_6^{18/7} (M/M_\odot)^{-5/7} \dot{m}_{17}^{-3/7} \text{ s} \quad (2)$$

where B_8 is the surface dipole magnetic field in units of 10^8 G, R_6 is the neutron star radius in units of 10^6 cm, M is its mass and $\dot{m}_{17}$ is the accretion rate in units of 10^{17} g s⁻¹. Thus, the weaker the neutron star's magnetic field, the shorter its period of rotation. However, the neutron star can only be spun up to the secular instability limit, given by equation (1) as

$$P_s = 1.05 \times 10^{-3} (M/M_\odot)^{-1/2} R_6^{3/2} \text{ s} \quad (3)$$

Equations (2) and (3) imply

$$B_8 \leq 1.91 (M/M_\odot)^{1/4} (R_6)^{-5/4} \dot{m}_{17}^{1/2} \quad (4)$$

Equation (2) is valid only if B is not so low that the Alfvén radius shrinks to the surface of the neutron star, requiring³

$$B_8 \geq 0.9 (M/M_\odot)^{1/4} R_6^{-5/4} \dot{m}_{17}^{1/2} \quad (5)$$

If the ram pressure of the infalling matter is substantial, then the dipole magnetic field structure near the neutron star's surface will possibly be distorted. In that case, equation (5) is not strictly valid; even so, it does provide a lower limit on B , if the simple disk accreting model for the galactic bulge sources is to be believed, and so is a plausible first estimate on how low B can be for fast pulsars. Substitution of $M_{\max}$ and $R_{\min}$ in equation (4) and of $M_{\min}$ and $R_{\max}$ in equation (5) then gives $B_{\max}$ and $B_{\min}$, the maximum and minimum values of the surface magnetic field, in terms of the parameter, $\dot{m}_{17}$.

After the process of accretion is over, the spun-up neutron star is expected to have a slow-down rate $\dot{P}$ of a freely rotating neutron star:

$$P^2/T \equiv P\dot{P} \sim \frac{8\pi^2 R_6^6 B^2}{3Ic^2} \quad (6)$$

(where T is its age) which gives, using equations (2) and (3),

$$\dot{P} \leq (2.25 \times 10^{-18}) (M/M_\odot) R_6^2 \dot{m}_{17} I_{45}^{-1} \text{ s s}^{-1} \quad (7)$$

where I_{45} is the moment of inertia in units of 10^{45} g cm².

Luminosities of galactic bulge X-ray sources are typically in the range 10^{37} – 10^{38} erg s⁻¹; so $\dot{m}_{17} \sim 1$ (W. Kundt, personal communication; see also ref. 3 and refs therein). The quantity $(M/M_\odot) R_6^2 I_{45}^{-1}$, for each EOS, is found to be a decreasing function of the central density of the neutron star and has a lower limit beyond which the hydrostatic equilibrium is not satisfied. The use of this lower limit in equation (7) gives an upper bound on $\dot{P}$. The calculated limiting values of $\dot{P}$ and B for the different EOS are shown in Table 2 (for $\dot{m}_{17} = 1$). The value of $\dot{P}$ measured for PSR1937+214 is 1.2×10^{-19} s s⁻¹. It lies within the predicted upper bounds on $\dot{P}$ for each EOS. From this alone, however, it is not possible to conclude anything regarding validity of any particular EOS at high densities.

Table 1 Bulk properties of rotating neutron stars at the secular instability limit for various EOS

EOS	$M_{\min}/M_\odot$	$R_{\max}$ (km)	$I_{\min}$ (g cm ²)	$M_{\max}/M_\odot$	$R_{\min}$ (km)	$I_{\max}$ (g cm ²)
Ref. 13	0.70	11.7	4.0×10^{44}	1.70	8.7	1.1×10^{45}
Ref. 14	0.76	12.0	5.0×10^{44}	2.08	9.2	1.6×10^{45}
Ref. 15 ($f^2 = 2.91$ case)	0.79	12.0	4.7×10^{44}	1.87	11.1	1.6×10^{45}
Ref. 16 (model I)	1.20	13.9	1.1×10^{45}	1.96	10.0	1.6×10^{45}
Tensor interaction ¹⁷	1.72	15.6	1.5×10^{45}	1.87	12.6	2.4×10^{45}

Values in columns 2–4 are as implied by PSR1937+214 and those in columns 5–7 are as given from the condition of hydrostatic equilibrium (see text).

Table 2 Theoretical limits on the surface magnetic field (B) and the slow-down rate ($\dot{P}$) for fast pulsars

EOS	$B_{\min}$ (G)	$B_{\max}$ (G)	$\dot{P}_{\max}$ (s s ⁻¹)
Ref. 13	3.80×10^6	1.46×10^7	2.92×10^{-18}
Ref. 14	3.76×10^6	1.43×10^7	2.36×10^{-18}
Ref. 15 ($f^2 = 2.91$ case)	3.80×10^6	1.10×10^7	3.06×10^{-18}
Ref. 16 (model I)	3.51×10^6	1.27×10^7	2.97×10^{-18}
Tensor interaction ¹⁷	3.32×10^6	9.43×10^6	3.66×10^{-18}
$\dot{P}_{\text{Obs}}(\text{PSR } 1937+214) = 1.2 \times 10^{-19} \text{ s s}^{-1}$			

The explanation for the low ($P, \dot{P}$) of fast pulsars as remnants of galactic bulge X-ray sources requires low surface magnetic fields for the neutron star. A value of $B \sim 3 \times 10^8$ G is inferred by Alpar *et al.*³, who took the neutron star parameters as: $M = 1M_{\odot}$, $I_{45} = 1$ and $\dot{m}_{17} = 1$. Using rotating neutron star models and realistic EOS, we find the maximum value of B for fast pulsars to be $\sim 10^7$ G. The implications of the low values of B are several. First, the reduction in the (anisotropic) opacity of photons will not be as drastic as in the case of the usual

pulsars¹⁰. Second, any angular momentum imparted to any part of the interior conducting (that is, electrically charged) components will not be promptly distributed throughout the pulsar, unlike the case for usual pulsars¹¹. Finally, because of low B , the cohesive energy at the surface¹² will be relatively small compared with the usual pulsars.

Received 23 November 1983; accepted 3 January 1984.

1. Backer, D. C., Kulkarni, S. R., Heiles, C., Davis, M. M. & Goss, W. M. *Nature* **300**, 615–618 (1982).
2. Radhakrishnan, V. & Srinivasan, G. *Curr. Sci.* **51**, 1096–1099 (1982).
3. Alpar, M. A., Cheng, A. F., Ruderman, M. A. & Shaham, J. *Nature* **300**, 728–730 (1983).
4. Fabian, A. C., Pringle, J. E., Verbunt, F. & Wade, R. A. *Nature* **301**, 222–223 (1983).
5. Tassoul, J. L. *Theory of Rotating Stars* (Princeton University Press, 1978).
6. Datta, B. & Ray, A. *Mon. Not. R. astr. Soc.* **204**, 75P (1983).
7. van den Heuvel, E. *Ann. N.Y. Acad. Sci.* **302**, 14 (1977).
8. Davidson, K. & Ostriker, J. P. *Astrophys. J.* **179**, 585–598 (1973).
9. Ghosh, P. & Lamb, F. K. *Astrophys. J.* **234**, 296–316 (1979).
10. Lodenquai, J., Canuto, V., Ruderman, M. A. & Tsuruta, S. *Astrophys. J.* **190**, 141–152 (1974).
11. Ruderman, M. A. *Rev. Astr. Astrophys.* **10**, 427–476 (1972).
12. Ruderman, M. A. & Sutherland, P. G. *Astrophys. J.* **196**, 51–72 (1975).
13. Pandharipande, V. R. & Bethe, H. A. *Phys. Rev. C* **7**, 1312 (1973).
14. Friedman, B. & Pandharipande, V. R. *Nucl. Phys. A* **361**, 502–520 (1981).
15. Canuto, V., Datta, B. & Kalman, G. *Astrophys. J.* **221**, 274–283 (1978).
16. Bethe, H. A. & Johnson, M. B. *Nucl. Phys. A* **230**, 1 (1974).
17. Pandharipande, V. R. & Smith, R. A. *Nucl. Phys. A* **237**, 507 (1975).

Role of surface reactions in the stabilization of n-CdS-based photoelectrochemical cells

Holly-Dee Rubin, Brian D. Humphrey & Andrew B. Bocarsly*

Department of Chemistry, Frick Laboratory, Princeton University, Princeton, New Jersey 08544, USA

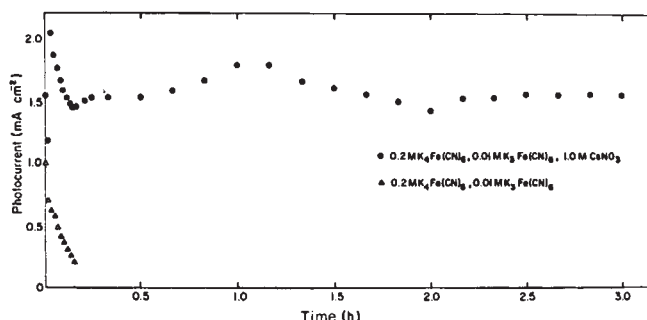
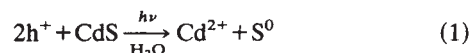


Fig. 1 Temporal behaviour of the output photocurrent for an n-CdS/Fe(CN)₆^{4-/-3-} photoelectrochemical cell potentiostated at -0.3 V versus SCE and containing either K⁺ only (Δ) or K⁺ and Cs⁺ (●). Both profiles are for cells irradiated with 40 mW cm⁻² of 488 nm light. A large surface platinum counterelectrode is used.

efficiency must be improved. If these problems are overcome, photoelectrochemical cells may become a viable alternative to traditional energy sources.

Although Gerischer⁶ demonstrated that addition of a K₄Fe(CN)₆/K₃Fe(CN)₆ mixture to the electrolyte of a n-CdS-based photoelectrochemical cell served to decrease slightly the rate of reaction (1), there is no clear understanding of the mechanism by which this occurs or how one might further enhance the CdS/Fe(CN)₆^{4-/-3-} cell stability.



The ability to suppress reaction (1) has previously been explained in terms of a kinetic competition between interfacial oxidation of solution Fe(CN)₆⁴⁻ and the rate of semiconductor photooxidation^{11,12}. The ability of the Fe(CN)₆^{4-/-3-} couple to compete successfully with the photodecomposition reaction has been attributed to the fact that the couple undergoes a moderately fast one-electron outer sphere charge transfer. However, other couples with redox kinetics and energetics similar to the Fe(CN)₆^{4-/-3-} couple do not increase cell stability¹³, and so the ability of Fe(CN)₆⁴⁻ to stabilize the n-CdS interface is to some extent anomalous.

Recent experiments carried out in our laboratory^{14,15} on the behaviour of oxidatively unstable metal electrodes in the presence of aqueous Fe(CN)₆⁴⁻ suggest that surface instability may be arrested by the reaction of near-surface metal ions with solution Fe(CN)₆^{4-/-3-} to form an electrode surface layer consisting of a [Mⁿ⁺Fe^{II/III}(CN)₆]^{n-4/n-3} lattice. Such surface layers

The potential use of II–VI semiconductor–aqueous junction cells for the conversion of optical energy to electricity has previously been limited by semiconductor photodecomposition processes combined with low energy conversion efficiencies. Decomposition processes in the prototypical n-CdS photoelectrochemical cell can be efficiently suppressed by addition of an appropriate polychalcogenide redox couple to the electrolyte^{1–3}. However, conversion efficiencies remain low (~5% at 488 nm)⁴. Moreover, although increased optical to electrical energy conversion rates can be obtained by using a redox couple such as Fe(CN)₆^{4-/-3-} (~8% conversion efficiency at 488 nm), the cell lifetime is greatly diminished^{5–7} ($t_{1/2} \sim 1/2$ h). We report here that the photodecomposition of n-CdS in a Fe(CN)₆^{4-/-3-} electrolyte can be dramatically decreased and cell output parameters significantly improved by the presence of an appropriate combination of K⁺ and Cs⁺ ions. Monochromatic (488 nm) conversion efficiencies in excess of 20% have been observed, with fill factors (a measure of the ideality of the cell) in the range of 65%. The enhanced stability and efficiency are associated with *in situ* chemical derivatization of the n-CdS surface with a layer of K_xCs_y[Cd^{II}Fe^{II}(CN)₆]. (This species is an analogue of Prussian blue having a C-bound Fe^{II/III} centre and a nitrogen bound Cd^{II} centre. See, for example, ref. 8.)

The growing interest in photoelectrochemistry has been stimulated by concern over the depletion of global supplies of fossil fuels. Photoelectrochemical cells provide one method for converting solar energy to electricity and chemical fuels (see ref. 9 for a discussion of how photoelectrochemical cells work). For example, inexpensive materials such as H₂O and CO₂ could be reduced, via a light-driven reaction, to such products as H₂ and CH₄. The major problem associated with such cells has been that of photocorrosion leading to decomposition of the semiconductor material. Another concern has been the low conversion efficiencies so far reported for these devices¹⁰. For liquid junction cells to be technologically feasible, stability and

* To whom reprint requests should be addressed.

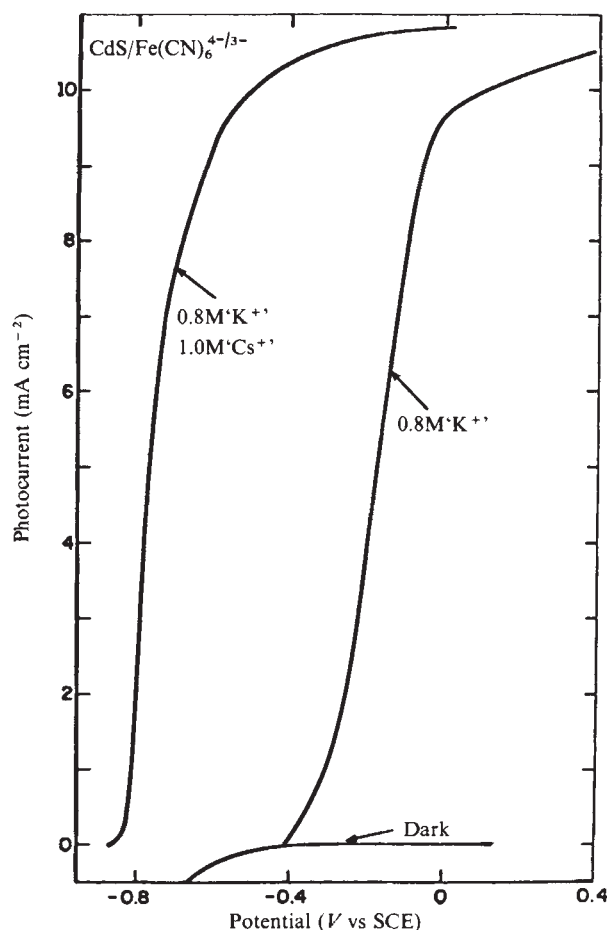


Fig. 2 The effect of Cs^+ on the steady-state current-potential response of a $\text{n-CdS/Fe(CN)}_6^{4-/3-}$ cell. The electrode is irradiated with 40 mW cm^{-2} of 488 nm light and the electrode potential is swept at 10 mV s^{-1} using a standard three-electrode potentiostat with a standard calomel reference electrode. The cell's dark response is not affected by Cs^+ . The ' K^+ -only' cell yields a fill factor of 35%, while the Cs^+ -containing system has a fill factor of 65%. For both cells the electrolyte redox potential is $+0.17 \text{ V}$ versus SCE.

can mediate charge transfer to solution species at enhanced rates. This, coupled with a physical protection mechanism, appears actively to inhibit electrode corrosion processes¹⁶. Thus, the photochemical generation of surface Cd^{2+} ions followed by reaction with solution $\text{Fe(CN)}_6^{4-/3-}$ to generate a surface film of covalently bound $[\text{Cd}^{\text{II}}\text{Fe}^{\text{II/III}}(\text{CN})_6]^{2-/1-}$ is likely to have a fundamental role in the observed stabilization process.

To test this hypothesis, both single crystal and polycrystalline pressed pellets of CdS were irradiated at open circuit (488 nm line of an Ar^+ laser) containing $0.2 \text{ M K}_4\text{Fe(CN)}_6$, $0.01 \text{ M K}_3\text{Fe(CN)}_6$ and 1 M KCl . The electrodes were then washed with water and analysed by diffuse reflectance Fourier transform IR spectroscopy. Surface spectra show the appearance of a band at $2,604 \text{ cm}^{-1}$, indicative of a bridging cyanide, and coinciding with the spectrum of an authentic sample of $\text{K}_2[\text{Cd}^{\text{II}}\text{Fe}^{\text{II}}(\text{CN})_6]$. In the case of single crystal samples the surface-attached species was only found on the Cd-rich (0001) face. Although the surface loading cannot be determined quantitatively by the IR techniques used, clearly greater than monolayer coverage was present¹⁵, ruling out simple chemisorption of Fe(CN)_6^{4-} and supporting the interfacial chemistry suggested above.

It is therefore possible that the unusual role of Fe(CN)_6^{4-} as a n-CdS antioxidant is due to its ability to generate an electrode surface capable of enhancing interfacial charge transfer rates via electron mediation, and not simply its heterogeneous charge transfer rate. To investigate this possibility further, investiga-

tions were carried out on both the redox behaviour of bulk $\text{K}_x[\text{Cd}^{\text{II}}\text{Fe}^{\text{II/III}}(\text{CN})_6]$ and the photoelectrochemical behaviour of derivatized n-CdS electrodes. These studies indicated that solid samples of $\text{K}_x[\text{Cd}^{\text{II}}\text{Fe}^{\text{III}}(\text{CN})_6]$ rapidly oxidize solution Fe(CN)_6^{4-} . Unfortunately, the rate of photoinitiated hole transfer from the n-CdS valence band to the surface-attached material is not rapid enough to preclude the formation of a hole population capable of carrying out reaction (1). Thus, although charge transfer mediation via surface $\text{K}_x[\text{Cd}^{\text{II}}\text{Fe}^{\text{II/III}}(\text{CN})_6]$ is both thermodynamically and kinetically possible, the rate of mediation is not sufficient to inhibit surface degradation processes totally.

This model can be further tested by taking advantage of the recent observations^{8,15,17,18} that Prussian blue analogues yield redox potentials which are strongly dependent on the alkali cations present. Thus, incorporation of appropriate cations in the surface layer could be expected to cause the redox potential of this species to become more positive, increasing the overlap between semiconductor valence band states and oxidizable $[\text{Cd}^{\text{II}}\text{Fe}^{\text{II}}(\text{CN})_6]^{2-}$ states, this latter effect increasing the rate of interfacial charge transfer. Experimentally we find that a combination of K^+ and Cs^+ performs this process. Addition of Cs^+ to the surface lattice^{8,17,19} is easily achieved by adding 1 M CsNO_3 to the electrolyte. Surface incorporation is possible since the surface lattice consists of a zeolitic structure²⁰ in which the migration of alkali cations is a low energy process which occurs rapidly when coupled to the iron centre redox reaction¹⁵.

As shown in Fig. 1, a spectacular enhancement in photocurrent stability occurs on this modification of the electrolyte. That this effect is not associated purely with a change in the electrolyte itself is indicated by the observations that the electrolyte redox potential is essentially unchanged ($\pm 15 \text{ mV}$) on addition of the CsNO_3 and that electrolytes containing either only Cs^+ or only K^+ do not enhance electrode stability, regardless of the concentration of the cation. Although these data do not provide evidence of total electrode stability, they clearly show that the $[\text{Cd}^{\text{II}}\text{Fe}^{\text{III}}(\text{CN})_6]^{2-/1-}$ surface derivative is a fundamental factor in the mechanism of stabilized interfacial charge transfer and that an appropriately constructed interface layer can dramatically affect the photostability of the n-CdS electrode.

The presence of Cs^+ also changes the cell's electrical output parameters. The major effect is a shift in the onset potential for photocurrent to values $\sim 400 \text{ mV}$ more negative than those observed when no Cs^+ is present (Fig. 2). Note that the maximum current obtainable is not affected by this process. In the best examples this shift in onset potential has yielded a cell with a maximum conversion efficiency under 488 nm irradiation of 22.6%. This efficiency is associated with a fill factor of 65% and a quantum yield for electron flow at the potential of the maximum power efficiency (-0.63 V versus SCE) of 0.74. More typically, as is the case for the data shown in Fig. 2, monochromatic cell efficiency is $\sim 15\text{--}18\%$, with associated quantum yields for electron flow of $\sim 0.5\text{--}0.6$. Thus, the overall effect is a $\sim 30\%$ increase in the fill factor coupled with a two- to threefold improvement in the maximum electrical output power $[(i \times V)_{\text{max}}]$ when compared with either Gerischer's results or a polysulphide-stabilized cell⁴. The shift in onset potential and the concomitant negative shift in the maximum power point allow the cell to be run at more negative potentials (without loss of efficiency) where a further enhancement in electrode stability is observed. The enhancement in cell output parameters can be linked with an increase in the ratio of interface charge transfer rates to all other excited state processes in low band bending conditions. This phenomenon is most easily explained if the highest occupied molecular orbital (HOMO) of the surface $\text{K}_x\text{Cs}_y[\text{Cd}^{\text{II}}\text{Fe}^{\text{II}}(\text{CN})_6]$ is considered to be a surface localized molecular orbital in good communication with the n-CdS valence band. (An alternative explanation for the increase in cell efficiency is that the surface derivative acts to remove or inhibit formation of a sulphur overlayer. Gerischer⁶ has previously indicated that n-CdS output parameters are sensitive to the presence of such a layer. Preliminary observations on the time evolution of the current-potential curve during surface

derivative formation suggests that this is not a likely explanation for the system of present interest.) Thus, even in minimal band bending conditions electron transfer from the surface material to the top of the valence band will be fast compared with nonproductive deactivation pathways or photodecomposition reactions.

This model is still under investigation in our laboratory; however, the picture would suggest that the exact cell efficiency is a strong function of the relative position of the surface species HOMO and the edge of the valence band. Hence, the $\text{CdX/Fe(CN)}_6^{4-/3-}$ ($\text{X} = \text{S, Se or Te}$) photochemical cell output parameters should be strongly affected by the exact cations present. Preliminary results based on a comparison of n-CdS and n-CdSe cells indicates this to be the case.

Received 10 October 1983; accepted 4 January 1984.

1. Ellis, A. B., Kaiser, S. W. & Wrighton, M. S. *J. Am. chem. Soc.* **98**, 1635–1637 (1976).
2. Hodes, G., Manassen, J. & Cahen, D. *Nature* **261**, 403–404 (1976).
3. Miller, B. & Heller, A. *Nature* **262**, 680–681 (1976).
4. Wrighton, M. S., Ellis, A. B. & Kaiser, S. W. *Proc. int. Symp. Solar Energy* (Electrochemical Society Inc., Princeton, New Jersey 1976).
5. Vanden Bergh, R. A. L., Gomes, W. P. & Cardon, F. *Z. phys. Chem. N. F.* **92**, 91–100 (1974).
6. Gerischer, H. & Gobrecht, J. *Ber. Bunsenges phys. Chem.* **80**, 327–330 (1976).
7. Gerischer, H. *J. electroanal. Chem.* **58**, 263–274 (1975).
8. Vol'khin, V. V. & Kalyuzhnyi, A. V. *Zh. prikl. Khim. Leningr.* **53**, 1216–1219 (1980).
9. Hammet, A. & Dennison, S. *Nature* **300**, 687–688 (1982).
10. Gronet, C. & Lewis, N. S. *Nature* **300**, 733–735 (1982).
11. Bard, A. J. & Wrighton, M. S. *J. electrochem. Soc.* **124**, 1706–1710 (1977).
12. Gerischer, H. *J. electroanal. Chem.* **82**, 133–143 (1977).
13. Kohl, P. A. & Bard, A. J. *J. Am. chem. Soc.* **99**, 7531–7539 (1977).
14. Bocarsly, A. B. & Sinha, S. *J. electroanal. Chem.* **137**, 157–162 (1982).
15. Sinha, S., Humphrey, B. D. & Bocarsly, A. B. *Inorg. Chem.* **23**, 203–212 (1984).
16. Bocarsly, A. B., Galvin, S. A. & Sinha, S. *Proc. Nickel Electrode Symp. Vol. 82*, 97–117 (The Electrochemical Society, Inc., Princeton, New Jersey, 1982).
17. Bocarsly, A. B. & Sinha, S. *J. electroanal. Chem.* **140**, 167–172 (1982).
18. Crumbliss, A. L., Lugg, P. S., Patel, D. L. & Morosoff, N. *Inorg. Chem.* **22**, 3541–3548 (1983).
19. Ellis, D., Eckhoff, M. & Neff, V. D. *J. phys. Chem.* **85**, 1225–1231 (1980).
20. Itaya, K., Ataka, T. & Toshimi, S. *J. Am. chem. Soc.* **104**, 4767–4772 (1982).

July–August temperature at Edinburgh between 1721 and 1975 from tree-ring density and width data

M. K. Hughes*, F. H. Schweingruber†, D. Cartwright* & P. M. Kelly‡

* Biology Department, Liverpool Polytechnic, Byrom St, Liverpool L3 3AF, UK

† Eidgenössische Anstalt für das Forstliche Versuchswesen, CH-8903 Birmensdorf ZH, Switzerland

‡ Climatic Research Unit, School of Environmental Sciences, University of East Anglia, Norwich NR4 7TJ, UK

The use of wood density measurements in dendroclimatology is a relatively recent development¹. There have been few attempts to use these data in climate reconstruction or to test the result rigorously. We now report on the reconstruction of July–August surface air temperature for Edinburgh, Scotland, for the period AD 1721–1975 using maximum latewood density and ring-width data from pine trees in the Scottish Highlands. In a series of tests, it is shown that the relationship between the climate and tree-ring variables is statistically significant and stable with time, confirming the power of dendroclimatic reconstruction even in a moist region such as the British Isles.

Tree-ring chronologies of ring width and intra-annual wood density variables were prepared for *Pinus sylvestris* L. from nine sites in northern Scotland using X-ray microdensitometry of increment cores². To allow for the effects of age on tree growth, chronologies of both ring widths and maximum latewood densities at each site were standardized by fitting a polynomial to each series and taking the quotient of the actual and fitted value for each year as the index. The individual core and tree index series were then averaged to produce a site chronology for ring width or density. These are standard techniques³. A new chronology parameter was used to assess the useful length

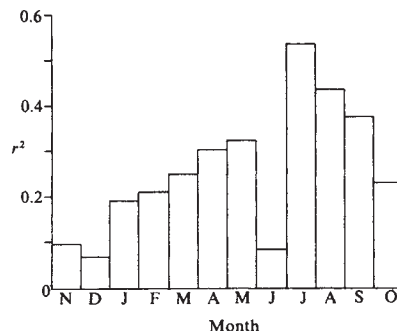


Fig. 1 Explained variance (r^2) for separate calibrations of tree-ring data on each month's mean temperature at Edinburgh for the years 1890–1969. Tree-ring data were index chronologies of ring width and maximum latewood density for current year i and years $i-1$ and $i+1$ at the five sites listed in Fig. 2 legend.

of a chronology as the number of cores decreases⁴ and the five longest pairs of site chronologies were used in this experiment.

After identifying the months in which the strongest link between Edinburgh temperature data⁵ and the tree-ring variables occurs, two calibration³ and verification^{3,6} exercises were undertaken. Calibration is the formalization of the climate–growth relationship in an equation, sometimes known as a transfer function, linking the climate and tree-ring variables and allowing reconstruction of the former from the latter. Verification is the testing of the reconstructed data by statistical comparison with independent climate data not used in calibration. The first exercise described here concerns the strength and stability of the relationship between climate and tree growth. The tree-ring variables were calibrated against 80 yr of the Edinburgh record (1810–89) and verified against an equal length of independent data (1890–1969). The calibration and verification periods were then reversed to check the stability of the results. In the second exercise, calibration was undertaken using 160 yr of the Edinburgh record (1810–1969) and the reconstruction based on this calibration was tested on Edinburgh data for the first 46 yr of the record (1764–1809) and on central England temperature data⁷, reduced to Edinburgh values, for an 89-yr period (1721–1809).

All calibrations reported here are based on a modification⁸ of the standard regression methods used in dendroclimatology³. Tree-ring data for the current year, i , and the previous and following years, $i-1$ and $i+1$, were presented as predictors to regressions in which Edinburgh temperature was the predictand. Data for years $i-1$ and $i+1$ were included to allow for autoregression in the tree-ring data and for lag effects³. The predictors were first screened using a stepwise multiple regression procedure in which values of F significant at $P=0.1$ and 0.15, respectively, were used to determine the entry and exit point of a predictor to and from the model. The principal components of the selected variables were then taken and the effective number of predictors further reduced by only retaining those components whose cumulative eigenvalue product exceeded unity. These components were entered into a final multiple regression producing the calibration equation in which Edinburgh temperature was expressed in terms of the tree-ring variables retained after the initial screening. Further details are given by Guiot⁸. A series of verification tests, described by Fritts³ and Gordon⁶, were performed on each reconstruction.

Although tree-ring width is believed to be related, through various mechanisms, to climate conditions over much of the year³, maximum latewood density has been found to be mostly a function of late summer temperature at high latitude and high altitude sites^{9–11}. Figure 1 shows the variance explained in calibrating the tree-ring width and density data on Edinburgh temperature data one month at a time. The strongest relationship is found in July and August; hence, the average temperature of these 2 months is selected for further analysis. The calibration for June is remarkably poor. This is consistent with the model

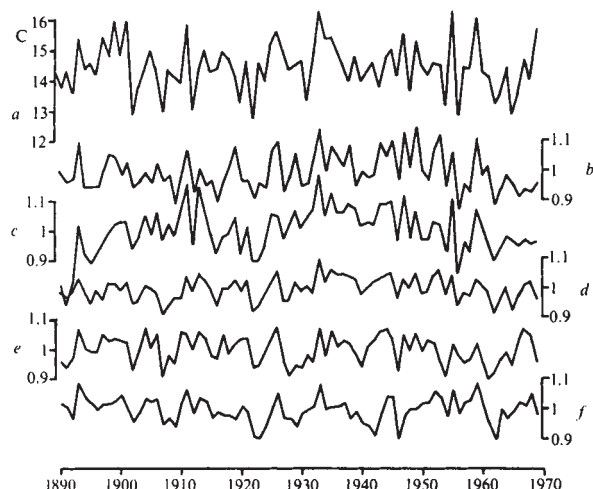


Fig. 2 Mean July–August temperature at Edinburgh, 1890–1969, compared with standardized maximum latewood density of *P. sylvestris* at five sites in the Scottish Highlands. *a*, Mean July–August temperature at Edinburgh; *b–f*, density chronologies. *b*, Ballochbuie, 56°57' N 3°19' W, 381 m above sea level (a.s.l.), north-east facing 15° slope, 11 trees sampled. *c*, Inverey, 57°0' N 3°35' W, 500 m a.s.l., west facing 15° slope, 12 trees. *d*, Glen Affric, 57°17' N 5°00' W, 300 m a.s.l., south-east facing 10° slope, 12 trees. *e*, Coulin, 57°32' N 5°21' W, 250 m a.s.l., north facing 45° slope, 10 trees. *f*, Loch Marree, 57°31' N 5°20' W, 100 m a.s.l., north-east facing 20°–45° slope, 8 Trees. The strong similarity between the density traces for the five sites is remarkable considering the heterogeneity of the sites. Although all are on relatively impoverished soils, there is a great variety of aspect and slope. Stand structure is, in most cases, open, with a preponderance of older trees, except at site *d* where the stand is denser. Sites *b* and *c* are close to the potential tree line, sites *d* and *e* are of 'intermediate status', whereas site *f* is well below the tree line. Conditions are strongly oceanic at sites *f* and *e* (mean annual rainfall ~2,000 mm) and decreasingly so from *d* to *c* and *b* (mean annual rainfall ~1,000 mm). A fuller account of the relationship between site and tree-ring measurements is being prepared.

for environmental control of maximum latewood density in conifers in cool, humid regions discussed by Schweingruber *et al.*¹², in which no secondary cell development occurs in a period in May and/or June. Possibly because maximum latewood density is dependent on one particular aspect of the seasonal climate, the year-to-year variations in all five chronologies of density data for the Scottish Highlands are similar despite a wide range of site, ecological and stand characteristics (Fig. 2). This, too, has been noted in other regions^{9,11}.

Three reconstructions of July–August temperature were undertaken (Fig. 3). The principal calibration and verification statistics^{2,6} are presented in Table 1. First, two 80-yr periods were used, in turn, for calibration and verification to test the stability of the relationship. Stability is essential if reconstruction is to be successful but instrumental records are rarely long enough for this to be adequately tested. Both calibrations account for a high proportion of predictand variance using relatively few of the 30 predictors. Only three predictors passed the initial screening before transformation into principal components and contribute to the calibration equation based on data for the period 1810–89 (density in year *i* at Inverey and Ballochbuie and ring width in year *i*+1 at Ballochbuie). Although more predictors contribute when the calibration periods are reversed, the major contributions come from the same three variables with similar coefficients and the same signs. In both cases, the verification tests indicate a high degree of similarity between the actual and reconstructed temperature data (see also Fig. 3a). The relationship between the tree-ring data and Edinburgh temperatures is both close and stable in time, and reconstruction is considered feasible.

Calibration was then undertaken for the 160-yr period 1810–1969 retaining 46 yr of actual temperature data for verification. The calibration statistics resemble those for the 80-yr calibrations and the serial correlation of the residuals is even closer to zero. Tree-ring variables from only three sites (density at Inverey and Coulin; ring-width and density at Ballochbuie) were selected by the regression procedure. The chronologies for these sites extend back to 1721 with adequate replication^{3,4}, so the temperature record was reconstructed back to this date using the 160-yr calibration. The verification statistics resulting from the comparison of the reconstructed data with actual Edinburgh temperature data for the period 1764–1809 are similar to those for the 80-yr period comparisons except that the reduction of error (RE) statistic is markedly smaller. As there is some doubt as to the reliability of the early temperature data (see Fig. 3 legend), a second set of tests was undertaken using central England temperature data reduced to Edinburgh values. The central England record extends back further than 1721, allowing verification over the full period of the reconstruction.

The series of Edinburgh July–August average temperatures was regressed on the central England data for the 160-yr period 1810–1969. As might be expected, the relationship is strong ($r=0.83$) and verification on the 46 yr of independent data indicated that the reduction was successful ($r=0.84$). Graphically, it can be seen that the actual data diverge from both the central England reconstruction and the tree-ring reconstruction about 1779 (Fig. 3b). This throws more doubt, on the grounds

Table 1 Calibration and verification statistics

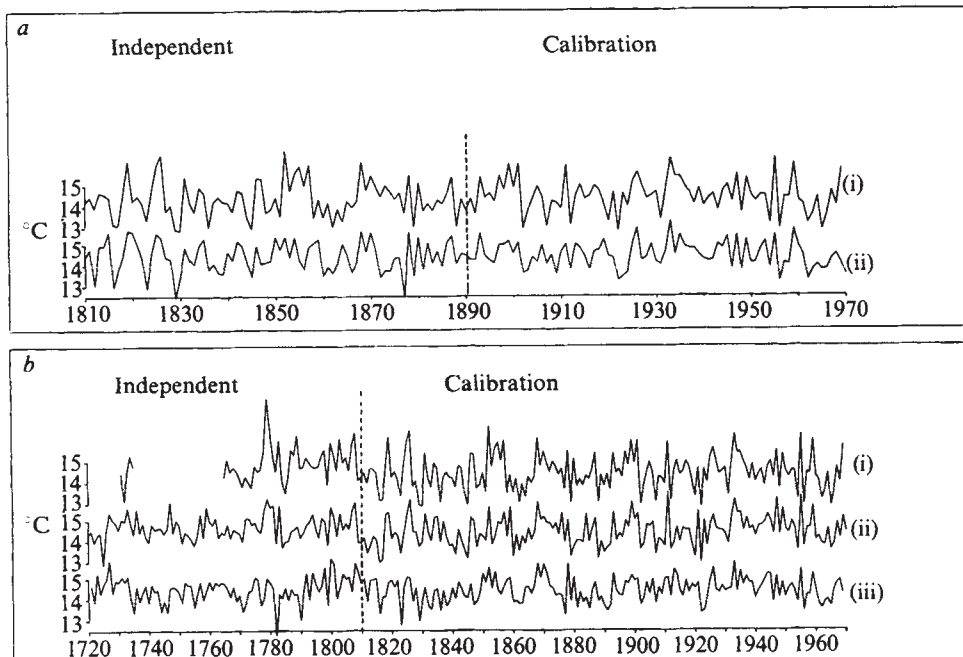
Calibration period	1890–1969	1810–1889	1810–1969	
Mean of actual values (°C)	14.46	14.39	14.42	
Standard deviation of actual values (°C)	0.83	0.88	0.85	
Multiple correlation coefficient	0.76‡	0.63‡	0.69‡	
Serial-correlation of residuals	–0.10	0.18	0.05	
No. of predictors retained	7	3	9	
Verification period	1810–1889	1890–1969	A	B
Mean of reconstructed values (°C)	14.41	14.41	14.50	14.49
Standard deviation of reconstructed values (°C)	0.74	0.46	0.62	0.70
Correlation coefficient	0.78‡	0.79‡	0.69‡	0.67‡
Reduction of error*	0.37	0.42	0.01	–0.12
Sign test (agreements/total)	63/79§	59/79§	31/45§	66/88§
Product means test ('t' value)†	4.75‡	4.82‡	2.73§	2.98§

* The reduction of error statistic is a measure of the residual error variance relative to that which we would obtain if the reconstructed values were compared with the mean for the calibration period. If $RE > 0$ there is some skill in the model compared with the trivial persistence estimate. The sign test involves a comparison of the signs of first differences in the two series.

† A significant positive product mean indicates a tendency for both actual and estimated departures from the average to be large when the sign is correctly estimated and to be small when the sign is incorrectly estimated: A, against actual Edinburgh record; B, against reconstruction of Edinburgh from central England record.

‡ $P < 0.001$; § $P < 0.01$.

Fig. 3 *a*, Edinburgh mean July–August temperature 1810–1969. (i) actual; (ii) reconstructed. The reconstruction shown is based on the calibration period 1890–1969. Thus, the reconstruction shown for 1810–89 is for an independent period. *b*, Edinburgh mean July–August temperature 1721–1969. (i) Instrumental record; (ii) indirect record derived from central England temperature record; (iii) tree-ring derived reconstruction using the calibration period 1810–69. The reconstruction shown for 1721–1809 is for an independent period. The instrumental Edinburgh temperature record extends back to 1764, but it is not based on observations from a single site. A detailed examination of its derivation reveals that it is based on records of decadal or greater length at Edinburgh back to 1784. More than one set of readings was available for most of the period back to 1785. From 1777 to 1783 the series is based on three different sites, two in Edinburgh lasting 3 yr and 18 months and one at Branhholm, 70 km SSE of Edinburgh, used for 2½ yr and reduced to Edinburgh values by reference to these last two short periods.



of consistency, on the 'actual' data for this period. The Edinburgh record is of complex origin at this time, and the 1779 value in particular, four standard deviations from the long-term mean and the most extreme departure in the record, must be considered suspect. Verifying the tree-ring reconstruction against the central England reconstruction, a high degree of similarity is indicated, again with the exception of a low RE statistic. This may be due to the small number of extremely

divergent values (for example, 1725) to which this statistic is particularly sensitive. The 5 yr of actual temperature data for the 1730s do not agree with either reconstruction and, again on the grounds of consistency, must be considered suspect.

Comparison of the variance spectra of the actual and reconstructed data (Fig. 4) indicates that the reconstruction has also been successful in the frequency domain. The redness of the spectra reflects the magnitude of the lag 1 serial correlation coefficient which is well reconstructed, being 0.11 in both series (after removal of linear trend). The slight disagreement at low frequencies (<0.1 cycles yr^{-1}) is not considered significant as the time series are too short to resolve spectral peaks with any certainty in this frequency band. The most notable similarity is the presence of the significant peaks centred at 2.13 yr in both spectra. These could well correspond to the quasibiennial oscillation observed in a number of Northern Hemisphere summer variables¹³. In view of their high temporal resolution, tree-ring measurements are the only proxy data capable of providing an extended record of variations on this time scale.

We have shown that maximum latewood density and ring-width measurements from pines in the Scottish Highlands can be used to reconstruct July–August temperature for Edinburgh, extending recent successes in European dendroclimatology^{9,14–16}. The feasibility of applying dendroclimatic methods in a moist region such as western Europe has been demonstrated strikingly. Light has been thrown on the reliability of one of the world's longest temperature records, and the success of the exhaustive verification of our reconstructions, only possible in an area such as western Europe where long instrumental records exist, has implications for the potential of the field of dendroclimatology as a whole. Further sampling of ancient living trees should allow the extension of the Edinburgh record back into the seventeenth century, the most severe phase of the Little Ice Age, and result in a valuable reconstruction. A proxy record for more distant times may be obtained using the large quantities of sub-fossil pine found in the British uplands.

Field collection was carried out by F.H.S. and D. C., density and ring-width measurements were made in F.H.S.'s laboratory, statistical analyses of the reconstructions were made by P.M.K. and all other analyses were undertaken in M.K.H.'s laboratory. We thank several colleagues for their helpful comments on an earlier draft of this paper.

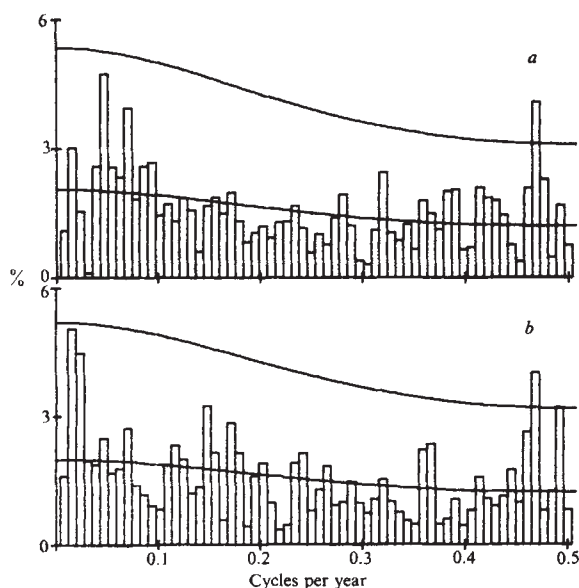


Fig. 4 Variance spectra of the reconstructed and actual data for the period 1764–1969. The lower line indicates the red noise null continuum and the upper line indicates the one-tailed 95% confidence limit. The spectral estimates were calculated using the fast Fourier transform (FFT) method following Rayner¹⁷. The raw data were detrended before analysis, tapered by a cosine bell and zeros were then added to bring the data length up to the power of two required by the FFT algorithm. The final spectral estimates are the averages of three separate sets based on independent sub-periods. *a*, Variance spectrum of mean July–August temperature at Edinburgh, 1764–1969. *b*, Variance spectrum of reconstructed July–August temperature at Edinburgh, 1764–1969, based on the 160-yr calibration period.

Received 14 October 1983; accepted 3 January 1984.

1. Schweingruber, F. H. in *Climate from Tree Rings* (eds Hughes, M. K., Kelly, P. M., Pilcher, J. R. & LaMarche, V. C.) 8–12 (Cambridge University Press, 1982).
2. Lenz, O., Schär, E. & Schweingruber, F. H. *Holzforschung* **30**, 114–123 (1976).
3. Fritts, H. C. *Tree Rings and Climate* (Academic, London, 1976).
4. Wigley, T. M. L., Briffa, K. R. & Jones, P. D. *J. Climate appl. Met.* (in the press).
5. Mossman, R. C. *Trans. R. Soc. Edinb.*, **38**, 681–755 (1896); **39**, 63–207 (1897); **40**, 406–509 (1902).
6. Gordon, G. in *Climate from Tree Rings* (eds Hughes, M. K., Kelly, P. M., Pilcher, J. R. & LaMarche, V. C.) 58–61 (Cambridge University Press, 1982).
7. Manley, G. *Q. J. R. met. Soc.* **100**, 389–405 (1974).
8. Guiot, J. *Institut d'Astronomie et de Géophysique, Université Catholique de Louvain Contr.* **21**, (1980).
9. Schweingruber, F. H., Fritts, H. C., Bräker, O. U., Drew, L. G. & Schär, E. *Tree-Ring Bull.* **38**, 61–91 (1978).
10. Parker, M. L. & Henoch, W. *Can. J. For. Res.* **1**, 90–98 (1971).
11. Conkey, L. in *Climate from Tree Rings* (eds Hughes, M. K., Kelly, P. M., Pilcher, J. R. & LaMarche, V. C.) (Cambridge, University Press, 1982).
12. Schweingruber, F. H., Bräker, O. U. & Schär, E. *Boreas* **8**, 427–452 (1979).
13. Landsberg, H. E., Mitchell, J. M., Crutcher, H. L. & Quinlan, F. T. *Mon. Weath. Rev.* **91**, 549–556 (1963).
14. Guiot, J., Berger, A. L. & Munaut, A. V. in *Climate from Tree Rings* (eds Hughes, M. K., Kelly, P. M., Pilcher, J. R. & LaMarche, V. C.) 160–163 (Cambridge, University Press, 1982).
15. Pilcher, J. R. & Hughes, M. K. in *Climatic Change in Later Prehistory* (ed. Harding, A. F.) (Edinburgh University Press, 1982).
16. Briffa, K. R., Jones, P. D., Wigley, T. M. L., Pilcher, J. R. & Baillie, M. G. L. *J. Climatol.* **3**, 233–242 (1983).
17. Rayner, J. N. *An Introduction to Spectral Analysis* (Pion House, Bristol, 1971).

Atmospheric P_{CO_2} changes recorded in lake sediments

D. Lal

Physical Research Laboratory, Ahmedabad 380 009, India and Scripps Institution of Oceanography, La Jolla, California 92093, USA

R. Revelle

University of California, San Diego, La Jolla, California 92093, USA

Recent studies of polar ice cores indicate that there have been appreciable secular variations^{1–4} in atmospheric CO_2 concentrations during the past 40,000 yr. We have now theoretically examined the expected steady state changes in atmospheric $\Delta^{14}C$ (a measure of the $^{14}C/^{12}C$ ratio) for different sizes of the atmospheric CO_2 reservoir, keeping fixed the rate of production of ^{14}C as well as the global carbon cycle exchange parameters. These calculations show that atmospheric $\Delta^{14}C$ is highly sensitive to changes in total atmospheric CO_2 . Appreciable changes in $\Delta^{14}C$ can also result from changes in the large-scale ocean circulation^{5–7}. Thus, the interpretation of the atmospheric $\Delta^{14}C$ palaeorecord requires supplementary geophysical evidence such as changes in sea level and in the carbonate chemistry of the oceans. We discuss the available data on past atmospheric $\Delta^{14}C$ with regard to changes in global carbon cycle parameters and in the rate of production of ^{14}C due to changes in Earth's dipole magnetic field.

Using a four-box model for the global carbon cycle, including the southern ocean reservoir, Lal and Venkatavaradan⁵ showed that changes in the atmospheric P_{CO_2} or in the large-scale ocean circulation can produce appreciable changes in the atmospheric $\Delta^{14}C$, of the order observed in the tree-ring record for the past 8,000 yr. The range is about 5–10% for plausible changes in these parameters. More recently, Siegenthaler *et al.*⁶ have come to similar conclusions using a box-diffusion model. The magnitude of atmospheric $\Delta^{14}C$ changes is not sensitive to the type of mixing model adopted, but it is sensitive to some of the exchange parameters. To bring out the key features of exchange, we present here theoretical relations for steady-state $^{14}C/^{12}C$ ratios in the atmosphere and in the mixed layer in terms of the carbon cycle exchange parameters for a three-box model, which is quite adequate for the present purposes.

We obtain the following relations for the specific activity of ^{14}C in the atmosphere and the mixed layer following the box

model formalism,^{5,8}

$$\frac{N_a^*/N_a}{N_d^*/N_d} = 1 + \frac{\lambda}{K_{ma}K_{dm}} [K_{ma} + K_{md} + K_{dm} + B + \lambda] \quad (1)$$

$$N_m^*/N_m = (N_d^*/N_d)(1 + \lambda/K_{dm}) \quad (2)$$

where N_i and N_i^* refer to stable carbon and radiocarbon contents of box i , K_{ij} refer to exchange coefficients between box i and j , subscripts a , m and d refer to atmosphere, mixed layer and deep ocean, respectively; B is the biological rate constant⁵ which determines the degree of imbalance between the total dissolved concentrations of CO_2 in the mixed layer, C_m , and in the deep sea, C_d ; λ is the disintegration constant of ^{14}C . The ratio of total concentrations of carbon in the deep sea and mixed layer is given by the following relation⁵:

$$C_d/C_m = 1 + B/K_{md} \quad (3)$$

Any changes in the size of the atmospheric reservoir will produce no appreciable changes in the specific activity of ^{14}C (N_d^*/N_d) in the deep sea and hence the ratio of specific activities in equation (1) can be considered to reflect primarily changes in N_a^*/N_a , the atmospheric specific activity of ^{14}C . If K_{ma} is kept constant, equation (1) predicts that the atmospheric specific activity of ^{14}C will hardly vary with any changes in the size of the atmospheric reservoir. However, if K_{am} is kept constant, K_{ma} would change with the size of the atmospheric reservoir as:

$$K_{ma} = \frac{N_a}{N_m} K_{am} \quad (4)$$

If K_{dm} were kept fixed, variations in N_a^*/N_a would arise due to changes in either the atmospheric inventory of carbon dioxide (via relation (4)) or biological productivity (via relation (3)), resulting in changes in B , K_{md} and K_{ma} .

Any variations in exchange parameters would affect the atmospheric $\Delta^{14}C$ values (relation (1)), but the most prominent effect comes from variation in K_{ma} as would be expected to occur due to changes in the atmospheric CO_2 inventory. Note that one would expect the air-sea CO_2 exchange flux to be proportional to the atmospheric P_{CO_2} .

The specific activity of the mixed layer would, however, remain unchanged with changes in B or atmospheric P_{CO_2} , as can be seen from relation (2). The pelagic remains in marine sediments cannot therefore provide any information on changes in the atmospheric CO_2 content.

As conventionally adopted, we take the mixed layer depth to be 75 m, the ocean depth 3,800 m and the amount of carbon in the present day atmosphere to be 58 m ocean equivalent. For K_{dm} and K_{am} , we similarly adopt working values of 10^{-3} and 10^{-1} yr^{-1} , respectively. Taking $C_d/C_m = 2.31/2.05$ based on GEOSECS data⁹ and with $K_{md} = 4.97 \times 10^{-2}$, corresponding to the value of $K_{dm} = 10^{-3}$, we find $B = 6.3 \times 10^{-3} \text{ yr}^{-1}$.

Results in Fig. 1 show appreciable changes in atmospheric $\Delta^{14}C$ with changes in atmospheric P_{CO_2} , the variation being asymmetric. The $\Delta^{14}C$ values change by +6.7 and -3.25%, respectively, for a factor of two decrease and increase in atmospheric P_{CO_2} . The result is in general agreement with the earlier calculations^{3,6}. Appreciable changes in atmospheric $\Delta^{14}C$ can also be caused by changes in the large-scale ocean circulation. Lal and Venkatavaradan⁵ showed that, in particular, changes in two of the parameters, the vertical mixing rate (K_{dm} and K_{md}) and the surface outcrop area of the Southern Ocean¹⁰, could easily produce changes in atmospheric $\Delta^{14}C$ of the order of 5–10%. These parameters are expected to vary with climatic changes^{5–7}. Secular variations in atmospheric $\Delta^{14}C$ have been modelled for transition from the last glacial to postglacial by Keir⁷ using a box model and by Siegenthaler *et al.*⁶ using a box-diffusion model. In these calculations, the authors have also considered changes in the ocean chemistry and the manner in which atmospheric P_{CO_2} may be expected to vary during the transition. We conclude that both changes in the atmospheric P_{CO_2} and global ocean circulation parameters can produce appreciable and comparable changes in atmospheric $\Delta^{14}C$, on

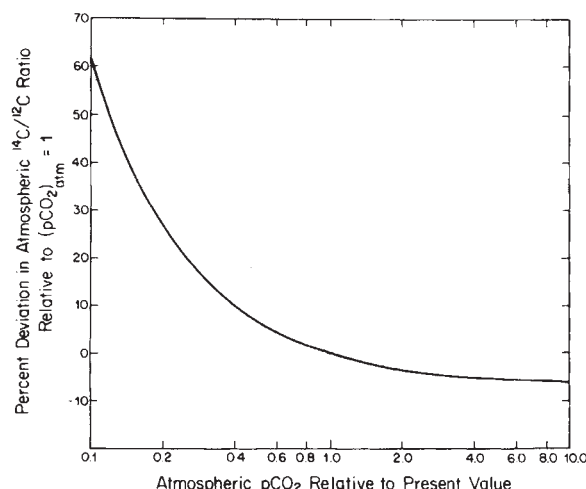


Fig. 1 Theoretically estimated deviations in atmospheric $^{14}\text{C}/^{12}\text{C}$ ratios for different atmospheric P_{CO_2} values relative to present P_{CO_2} value (=1). The calculations are based on a fixed residence time of CO_2 in the atmosphere, 10 yr.

time scales 100–10,000 yr, depending on the manner in which ocean chemistry and circulation changed with climate.

Figure 1 predicts that the atmospheric $\Delta^{14}\text{C}$ value would have been higher by about 7% when atmospheric P_{CO_2} was $\sim 50\%$ of present value, about $15\text{--}20 \times 10^3$ yr BP. Unfortunately tree-ring data exist only as far back as 8,000 BP¹¹. We must confine ourselves to continental sediment data. We are aware only of the relevant data published by Stuiver¹² for the sediments from the Lake Jih Tan which go back to about 35,000 BP. The $^{14}\text{C}/^{12}\text{C}$ activity ratio (normalized to unity at the sediment surface) is plotted in Fig. 2 as a function of depth; the depth of a layer is expressed in terms of cumulative mass of non-combustible sediment above the layer. It appears that the simplest assumption one can make to determine relative variations in the $^{14}\text{C}/^{12}\text{C}$ activity ratios is to assume constant sedimentation rates, denoted as S , for the 0–70 and 70–145 g cm^{-2} depth intervals. The corresponding $^{14}\text{C}/^{12}\text{C}$ ratios at deposition oscillate around the mean with an amplitude smaller than 10% except for the depth interval 88–98 g cm^{-2} , where one observes a marked positive excursion of peak magnitude 20%. Based on the sedimentation model, one deduces an age bracket of $15\text{--}20 \times 10^3$ yr BP for this layer. Clearly, in the absence of a reliable varve chronology, the lake sediment data must be treated with caution. However, any prominent short duration excursions such as the one noted above would probably be due to changes in atmospheric $^{14}\text{C}/^{12}\text{C}$ ratios.

The average positive excursion of 10% in $^{14}\text{C}/^{12}\text{C}$ ratios at $15\text{--}20 \times 10^3$ yr BP (88–98 g cm^{-2} depth; Fig. 2), if attributed solely to changes in atmospheric P_{CO_2} , corresponds to about 60% lower atmospheric P_{CO_2} than at present (Fig. 1). Of course, this calculation is most sensitive to the value of K_{atm} , for which we have adopted a value of 0.1 yr^{-1} , corresponding to a residence time of 10 yr for CO_2 in the atmosphere. Based on studies of bomb ^{14}C in banded corals, Druffel and Suess¹³ have obtained values of 16–30 yr for the residence time of CO_2 ; a value of 20 yr may be more realistic. However, that part of the observed change in the atmospheric $^{14}\text{C}/^{12}\text{C}$ ratio at $15\text{--}20 \times 10^3$ yr BP may have been due to a change in the vertical mixing rate of the ocean.

Thus, we have shown theoretically that atmospheric $^{14}\text{C}/^{12}\text{C}$ ratios vary inversely with concentration of CO_2 in the atmosphere. This result follows from the reasonable assumption that the flux of CO_2 from the atmosphere to the sea is proportional to the amount of CO_2 in the atmosphere. This conclusion suggests a new means of determining past variations in atmospheric CO_2 provided that reliable palaeodata can be obtained on atmospheric $^{14}\text{C}/^{12}\text{C}$ ratios. Unfortunately, other causes such as variations in production rate of ^{14}C and ocean circulation pattern can also produce appreciable variations in atmospheric

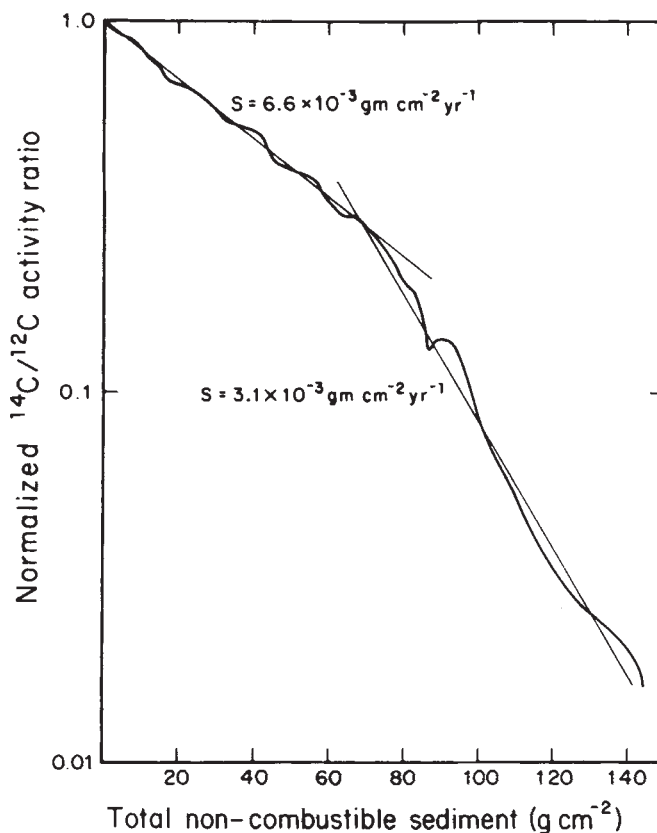


Fig. 2 Measured $^{14}\text{C}/^{12}\text{C}$ activity ratios in Lake Jih Tan (Taiwan) normalized to unity at surface, as a function of sediment depth. Original data are due to Stuiver¹². The two constant sedimentation rate lines intersect at 71 g cm^{-2} corresponding to 10,780 yr BP. Note the large excursion in $^{14}\text{C}/^{12}\text{C}$ ratio at 88–98 g cm^{-2} depth, which corresponds to $15\text{--}20 \times 10^3$ yr BP.

^{14}C specific activity. Therefore, to ascertain the contributions of changes in atmospheric P_{CO_2} to any observed variations in atmospheric $\Delta^{14}\text{C}$, one would have to call on supplementary information which bears on geomagnetic data (to deduce changes in global production rate of ^{14}C) and on ocean circulation parameters.

Analyses of CO_2 in ice cores from Greenland and Antarctica indicate that during part of this time, at $15\text{--}20 \times 10^3$ yr BP, the atmospheric CO_2 was only half the early twentieth century value¹⁴. The effect of this should be easily seen in the atmospheric $\Delta^{14}\text{C}$ palaeorecord. Tree-ring data go as far back as 8,000 yr BP but the possibility exists that lake sediments, which should closely reflect the atmospheric $^{14}\text{C}/^{12}\text{C}$ ratios¹⁴, might be used. One series of $^{14}\text{C}/^{12}\text{C}$ measurements has been published¹² on lake sediments deposited during the latter part of the last glacial epoch $10\text{--}30 \times 10^3$ yr BP. A clear-cut increase of the expected magnitude in $^{14}\text{C}/^{12}\text{C}$ ratios is observed in the lake data (Figs 2, 3). This result gives credence to the evidence³ from ice cores that both atmospheric CO_2 and temperature (indicated by $^{18}\text{O}/^{16}\text{O}$ ratios) were lower $15\text{--}20 \times 10^3$ yr ago than either before or since.

So far, we have treated the problem of atmospheric P_{CO_2} changes without being concerned about its causative mechanisms, that is, changes in the marine carbonate chemistry. But what caused the CO_2 change? Broecker¹⁵ and McElroy¹⁶ have ascribed this to a variation in biological productivity of the oceans resulting from a change in the amount of dissolved phosphorus and fixed nitrogen, respectively. When these nutrients are relatively abundant, biological productivity is high; more CO_2 is transferred from surface to deep waters by the biological pump—the gravitational sinking of organic matter from near surface waters.

The principal difficulty with the above hypotheses arises from the fact that because of the long residence times of total oceanic

P and N, one would expect changes in P_{CO_2} to occur over long periods of time, several thousands of years or more. The observations of Stauffer *et al.*³, however, show that during glacial epochs, marked changes in atmospheric P_{CO_2} commonly took place over periods as short as a few hundred years.

We believe these short-term variations are genuine and expected to arise from small perturbations in oceanic circulation in the top 1 km depth. The oceanic control on atmospheric P_{CO_2} on inter-annual and decadal time scales is borne out by recent observations. Changes in P_{CO_2} in near surface waters corresponding to several tens of parts per million have occurred over periods of the order of a decade, at least at high latitudes^{17,18}. Bacastow *et al.*¹⁹ have reported anomalous rates of CO_2 increase in the atmosphere inversely correlated with the Southern Oscillation Index. Gammon and Komhyr²⁰ have shown that the observed slower CO_2 growth in the atmosphere in 1982, ~50%, is a combined effect of weakened equatorial Pacific Ocean source and enhancement of ocean sink for CO_2 at high latitudes (>40°) due to large negative sea-surface temperature anomalies (1–3 °C) beginning in mid-1982, particularly in the intermediate and deep water formation regions of North Atlantic and Pacific oceans. The observed high degree of correlation between changes in CO_2 and $^{18}O/^{16}O$ ratios in ice³ demonstrates the importance of positive feedback mechanism of changes of P_{CO_2} in the atmosphere. Thus, small changes in the atmospheric P_{CO_2} due to changes in circulation in the top 1 km of the oceans, particularly at high latitudes, may be greatly amplified to lead within a few decades to large decreases or increases in atmospheric P_{CO_2} and temperature.

Finally, concerning the generally accepted belief that atmospheric $\Delta^{14}C$ palaeorecord for the past 8,000 yr supports the large change in the Earth's dipole field based on archaeomagnetic data²¹, the record shows two prominent features¹¹: a high frequency variation with a period of about 200 yr and a slow variation with a period of 8,000 yr. The former is clearly due to solar modulation of galactic cosmic ray flux²² (which results in a variation in global production rate of ^{14}C). The latter is believed to be due to change in the Earth's dipole field^{11,21}. The present work highlights the fact that a part or most of slow variation may in fact have been due to changes in the atmospheric P_{CO_2} and in ocean circulation. Combining the P_{CO_2} record in ice cores with the tree-ring and lake sediment $\Delta^{14}C$ record, it should be possible to delineate changes occurring over different time scales in the oceanic circulation/chemistry and the Earth's dipole field.

We thank W. S. Broecker and M. Stuiver for useful comments, R. S. Keir for helpful discussions, and H. Oeschger and U. Siegenthaler for comments on the manuscript.

Received 27 June 1983; accepted 6 January 1984.

- Berner, W., Oeschger, H. & Stauffer, B. *Radiocarbon* **22**, 227–235 (1980).
- Delmas, R. J., Ascencio, J. M. & Legrand, M. *Nature* **284**, 155–157 (1980).
- Stauffer, B., Hofer, H., Oeschger, H., Schwander, J. & Siegenthaler, U. *Proc. Symp. Ice and Climate Modelling*, Evanston, Illinois; preprint *Annls Glaciol.* (in the press).
- Lorius, C. & Raynaud, D. preprint, Laboratoire de glaciologie et de géophysique de l'environnement, Grenoble (1983).
- Lal, D. & Venkatavaradan, V. S. in *Radiocarbon Variations and Absolute Chronology* (ed. Olsson, I.U.) 549–569 (Wiley, New York, 1970).
- Siegenthaler, U., Heimann, M. & Oeschger, H. *Radiocarbon* **22**, 177–191 (1980).
- Kier, R. S. *Earth planet. Sci. Lett.* **64**, 445–456 (1983).
- Craig, H. *Tellus* **9**, 1–17 (1957).
- Takahashi, T., Broecker, W. S. & Bainbridge, A. E. in *Carbon Cycle Modelling: Scope 16* (ed. Bolin, B.) 271–286 (1981).
- Craig, H. in *Earth Sciences and Meteoritics* (eds Geiss, J. & Goldberg, E. D.) 103–114 (North-Holland, Amsterdam, 1963).
- Suess, H. E. *Radiocarbon* **22**, 200–209 (1980).
- Stuiver, M. in *Radiocarbon Variations and Absolute Chronology* (ed. Olsson, I.U.) 197–213 (Wiley, New York 1970).
- Druffel, E. M. & Suess, H. E. *J. geophys. Res.* **88**, 1271–1280 (1983).
- Broecker, W. S., Peng, T.-H., Mathieu, G., Hessler, R. & Torgersen, R. *Radiocarbon* **22**, 676–683 (1980).
- Broecker, W. S. *Geochim. cosmochim. Acta* **46**, 1689–1705 (1982).
- McElroy, M. B. *Nature* **302**, 328–329 (1983).
- Keeling, C. D. *J. geophys. Res.* **73**, 4543–4553 (1968).
- Takahashi, T., Chipman, D. & Volk, T. *Proc. Carbon Dioxide Research Conf.* II-123-145 (CONF-820970, Institute of Energy Analysis, Washington DC, 1983).
- Bacastow, R. B. *et al. Science* **210**, 66–68 (1980).
- Gammon, R. H. & Komhyr, W. D. in *IUGG 1983 Hamburg Programme and Abstracts*, **2**, 828 (International Union of Geodesy and Geophysics, 1983).
- Bucha, V. in *Radiocarbon Variations and Absolute Chronology* (ed. Olsson, I.U.) 501–511 (Wiley Interscience, New York, 1970).
- Castagnoli, G. & Lal, D. *Radiocarbon* **22**, 133–158 (1980).

Ocean-wide stagnation episode in the late Cretaceous

P. C. de Graciansky*, G. Deroo†, J. P. Herbin†, L. Montadert†, C. Müller†, A. Schaaf‡ & J. Sigal§

* Ecole Nationale Supérieure des Mines de Paris, Centre de Géologie Générale et Minière, 60 bd Saint Michel, 75272 Paris Cédex 06, France

† Institut Français de Pétrole, 1 et 4 avenue de Bois Préau, B.P. 311, 92506 Reuil-Malmaison Cédex, France

‡ Institut de Géologie, 1 rue Blessig, 67084 Strasbourg Cédex, France

§ 33 rue de Montreuil, 94300 Vincennes, France

In this paper we report that marine sediment locations spread over a wide area contain evidence of an anoxic interval during the mid-Cretaceous, here called the Cenomanian–Turonian Black Shale Horizon (CTBSH). The occurrence of laminated carbonaceous shales with marine organic matter in an unoxidized state implies deposition and early diagenesis in anaerobic environments. The stratigraphic record above and below the CTBSH in environments as different as the deep North Atlantic Ocean and marginal and epicontinental seas over enormous areas in North America, North Africa and northwestern Europe is characterized by a stratigraphic gap, a disconformity or a continuous but strongly condensed stratigraphic section. The anoxic layer coincides in time with the maximum extension of a worldwide marine transgression of eustatic origin recorded on continental margins¹ as well as on various cratons². No simple model has yet been found to explain the occurrence of the CTBSH in such a wide range of depth and palaeoenvironmental conditions.

Our revision³ of the stratigraphy and organic geochemistry of the North Atlantic DSDP boreholes has demonstrated the CTBSH at Sites 105, 137, 367, 386, 398, 549 and 551 (Figs 1 and 2). The sediment is usually pyritic, phosphatic or both, and is laminated on a millimetre or submillimetre scale. The total organic carbon (TOC) content is more than 2% and usually reaches 10% on the average and more than 30% at its maximum, with organic matter mainly of marine origin²².

These holes suggest three different modes of occurrences of the CTBSH determined by the evolution of redox conditions in the bottom sea water. Thus at DSDP Site 105 (North American Lower Continental Rise Hills), the CTBSH represents an episode of marked anoxia and slow sediment deposition ending a period of alternating oxia/anoxia and followed by deposition in a well oxygenated environment. The CTBSH is 1-m thick with maximum TOC of 24%. It overlies alternate light-coloured (TOC < 0.5%) and black (TOC < 5%) laminated claystones of Albian and early Cenomanian age (Hatteras Formation) and is overlain by reddish brown claystones (Planagenet Formation). The accumulation rate for the CTBSH is 0.25 to 0.2 m Myr⁻¹, compared with 16 to 7 m Myr⁻¹ in the Albian to the early Cenomanian and less than 1 m Myr⁻¹ in the Senonian. By contrast, at DSDP Sites 549 and 551⁴, located at mid-depth on the slope of the Celtic margin in the North-East Atlantic, the CTBSH was deposited during a short period of anoxia within a prolonged oxygenated interval as recorded from neighbouring epeiric seas⁵. Maximum TOC is near 10%. The anoxic sediments are interbedded between chalks lean in organic carbon. Again, the accumulation rate was less than 1 m Myr⁻¹. Palaeodepths are estimated at approximately 1,500 m at both sites.

At DSDP Site 367 (Cape Verde Basin, West Africa), late Aptian to early Cenomanian laminated siltstones and mudstones were accumulated at average rate 14 m Myr⁻¹, with maximum TOC of 30% and average TOC of 10%. The CTBSH caps these deposits (core 10 of site 367) with the highest TOC values (31 to 39%) measured in the Cretaceous of the North Atlantic. The

Fig. 1 Minimum geographic extension of the Cenomanian–Turonian black shale horizon in the North Atlantic, northwestern Europe and the Tethys. (1) Extension of the CTBSH; (2) extension of late Cretaceous deposits removed by subsequent erosion; (3) presence of the CTBSH at DSDP wells; (4) hiatus dated Cenomanian–Turonian at DSDP wells; (5) well ending within oceanic basalts.

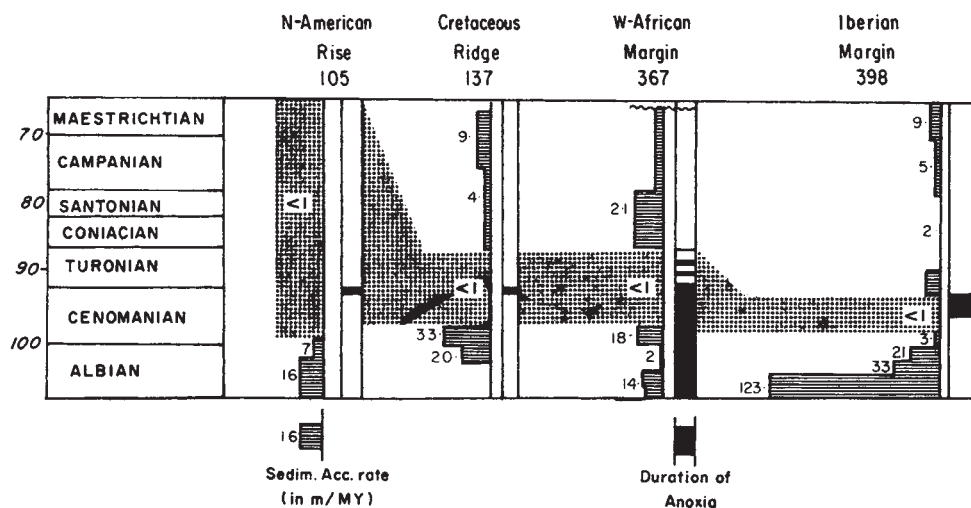
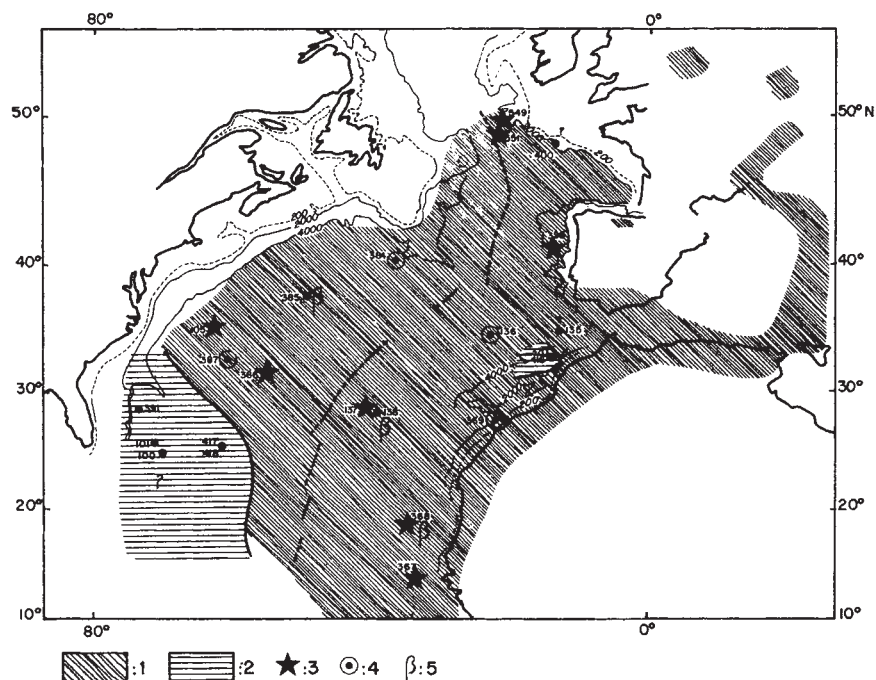


Fig. 2 Estimates of sediment accumulation rate and duration of anoxia in selected DSDP bore holes. This diagram shows the episode of anoxia and of sedimentary condensation in the late Cenomanian–Turonian. Stipple indicate the period of accumulation rate less than 1 m per Myr. Ages were determined from biostratigraphical data.

sedimentation rate is less than 1 m Myr^{-1} . The deposition of the CTBSH thus ends a period of 15 Myr of permanent anoxia. Thereafter oxygenation of the basin took place progressively and rhythmically during the early Senonian, as shown by alternating light-coloured and carbonaceous dark mudstones with TOC values oscillating between 0.5% and 10–15%. This site shows the typifying characteristics of the CTBSH—a duration of $\approx 12 \text{ Myr}$, the importance of slow deposition (0.8 m Myr^{-1}) and the high degree of anoxia³.

Comparison of the DSDP data from Site 370³ field observations on the adjacent Moroccan landmass^{6–9} show that in the basins located at the outer part of the West African margin, the anoxia lasted for the long interval between the Albian and the Coniacian, but that the duration of the anoxia becomes progressively less on the outer shelf and is restricted to only the late Cenomanian–early Turonian in the epicontinental northwestern African domains (Fig. 3), where it coincides with the maximum of the major early Turonian eustatic transgression². In the Tarfaya and Agadir basins, the outermost basins of the Moroccan margin in the Cretaceous, bituminous shales deposited at an estimated palaeodepth of 300 m show a relatively wide stratigraphic extent, from late Cenomanian to Coniacian^{6,23,24}, comparable with that observed at deeper Atlantic sites, for example DSDP hole 370. The inner margins and the epeiric domain which constituted the Cretaceous “Atlas Gulf” is now represen-

ted by the Meseta basin and in the High Atlas folding, where bituminous facies with high TOC (22%) in the carbonate-free sediments are everywhere found at the bottom of a massive carbonate sequence of late Cenomanian to Turonian age (*barre turonienne*) over a wide region. Farther inland, in the Sahara Cretaceous basins, there is a major change in sedimentation from Cenomanian gypsiferous red beds to Turonian white and grey rock colours.

Jenkyns⁷ has shown that the CTBSH also occurs on other peri-Atlantic continental margins such as Venezuela, Trinidad, the Caribbean and in various epeiric seas, the North Sea, South-East England⁵, western Germany, Nigeria and the western interior of the United States. Extension of the CTBSH is found in the Tethyan domains in palaeoenvironments, ranging from the deep ocean to epeiric seas, as varied as in the Atlantic.

In the North African coastal ranges, the palaeotransform zone connected the central Atlantic Ocean to the Piedmontese-Ligurian micro ocean was marked by a narrow trough in which thick detrital graded sequences (flyschs) were deposited during the Cretaceous. Carbonaceous cherts (phtanites) have long been known in Chouamat-Melloussa flysch units in Algeria and have now been dated as late Cenomanian and early Turonian^{8–12}. They continue into Sicily as the Troina Series, which comprise part of the major Reitano Nappe where black Cenomanian–Turonian bituminous horizons contain fish debris and oysters.

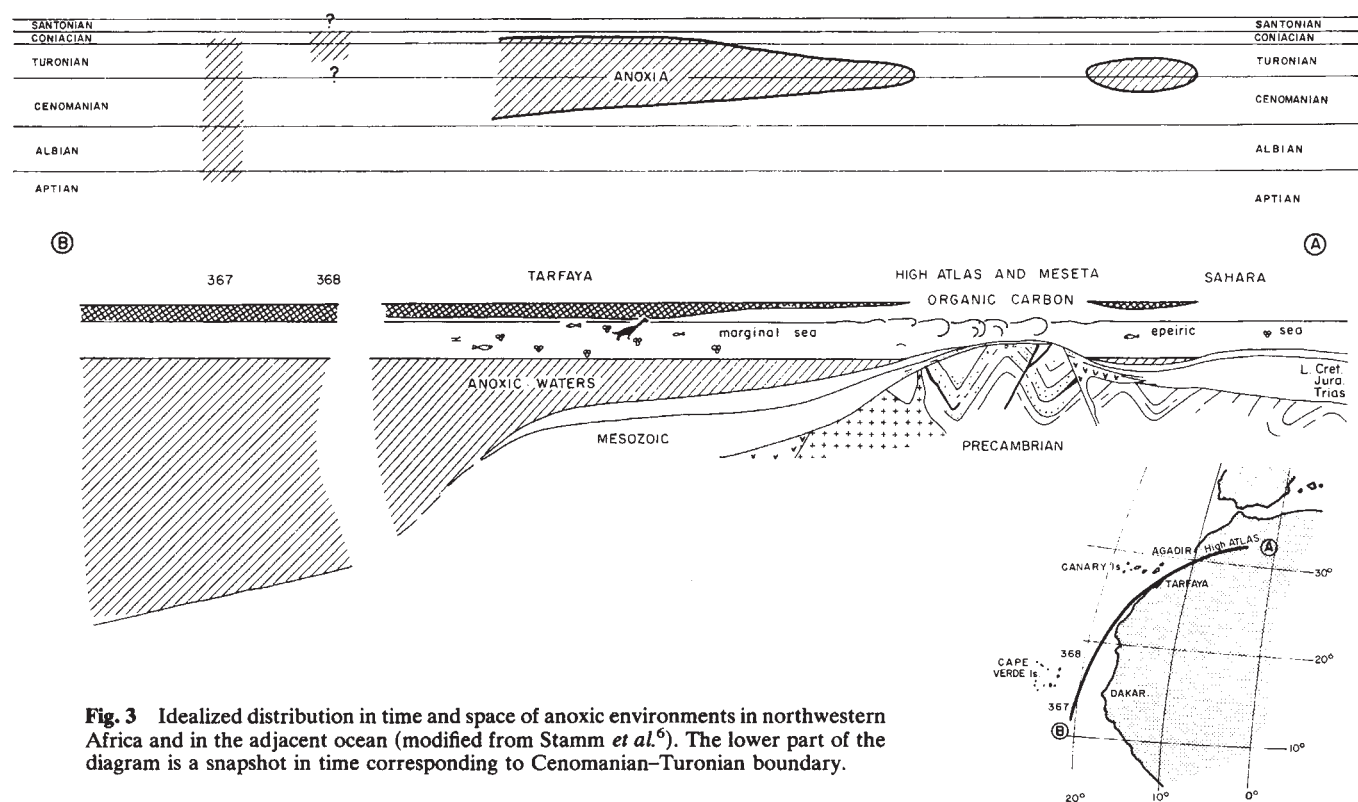


Fig. 3 Idealized distribution in time and space of anoxic environments in northwestern Africa and in the adjacent ocean (modified from Stamm *et al.*⁶). The lower part of the diagram is a snapshot in time corresponding to Cenomanian–Turonian boundary.

Whereas the flysch sequences are characterized by a high accumulation rate, the phanites corresponding to the CTBSH are marked by a temporary low sedimentation rate comparable with that in the Atlantic realm.

The Carbonaceous black-shale interval can also be followed southwards in the deeper parts of epeiric Cretaceous seas which have emerged in the northern Sahara, Algeria⁹ and central Tunisia¹³ and can be reconstructed from very clear field descriptions scattered in the *Publications du Service de la Carte Géologique d'Algérie* dated 1957–74.

Most of the Piedmontese-Ligurian small palaeo-ocean is now contained within a thick pile of thrust units extending over much of the Apennines and the Italian, French and Swiss Alps. Thrust sheets originating from the deeper parts of the micro-ocean are severely deformed if not metamorphosed, so that stratigraphic correlation of the black shale intervals with the Atlantic CTBSH is uncertain⁷. On the other hand autochthonous (Apulian block) and para-autochthonous domains (Marche, Umbrian, Campanian) derived from marginal and epicontinental units of the Tethys demonstrate clear equivalence to the CTBSH. There, foraminiferal and nannofossiliferous calcareous mudstones constituting the Scaglia are interbedded with a narrow layer of bituminous black shale known as *Livello Bonarelli* which contains numerous fish remains and has a TOC as high as 13%^{7,14}. Similar lithologies have been identified as far away as the southern Alps, in sections of the *Scaglia Variegata* in the Trento zone¹⁵. The rate of deposition of the entire Bonarelli horizon is estimated between 0.35 and 0.7 Myr¹⁶.

The geographic extent and short duration of the CTBSH both attest the importance and the specific nature of this episode of marine anoxia in the Cretaceous. Indeed the CTBSH is most clearly evident in the central North Atlantic and the western Tethys, probably has an equivalent as far away as the Mariana basin in the West-Pacific which was recently drilled at DSDP site 585¹⁷.

The preservation or the destruction of organic matter on the sea bottom depends essentially on the oxygen concentration in the bottom sea water, itself determined by the dynamic equilibrium between oxygen consumption, mainly by decay of organic matter, and renewal by bottom currents carrying oxygenated waters. Modest organic productivity in almost completely

restricted waters, as in the Black Sea, leads to the preservation of organic matter in the sediments, as does high organic productivity that outsteps oxygen replenishment by bottom currents^{6,18}. The model of the upwelling has been proposed to explain the Cenomanian carbonaceous shales of Morocco⁶ or of northeast Atlantic¹⁸. However, the wide geographical extent as well as the different palaeodepths of deposition of the CTBSH from the American Western Interior basins to the Tethys argues against the applicability of this concept as the only factor.

The known occurrences of the CTBSH correspond to a relatively wide range of organic productivity. The deposition of the CTBSH coincided with a period of relatively high productivity is supported by the development of radiolarian cherts at this level; these have been observed at DSDP sites 137, 398, 417, 549 and 551 in the western and northeastern Atlantic, on the Moroccan craton⁶ and in many outcrops of the Tethyan units and of the Apennines. The cherts are associated with sporadic lava flows in the Massylian units of Algeria¹⁹ and with volcanic ash in the North Sea neighbourhood⁹ which could have contributed locally to the siliceous biogenic productivity.

However, it seems that the productivity was relatively moderate in many other parts of the Atlantic. At the foot of the West African margin, the high TOC content is not associated with siliceous biogenic layers. Zeolites and opal-CT which generally derive from the tests of radiolarians are usually much less abundant than in the Western part of the Atlantic³. Most probably the West African sub-basin conforms with the euxinic (Black Sea) model. In the shallow Cretaceous sedimentary environments of northwestern Europe⁵, Morocco⁶, and in the southern Alps (Trento plateau¹⁵) the palaeomorphology of the seabed may have been influential in trapping anoxic waters or in allowing superficial re-oxygenation (Fig. 3).

The stratigraphic record around the Cenomanian–Turonian boundary in marginal or epicontinental palaeoenvironments in North America, Africa and Europe is everywhere characterized by either (1) a stratigraphic gap or even a disconformity or (2) a continuous but markedly condensed stratigraphic section. These characteristics are found even in many flysch sequences formed by rapid accumulation while in several cases for example in South-East England⁵ they are found together. These features may simply be related to the early Turonian high stand of sea

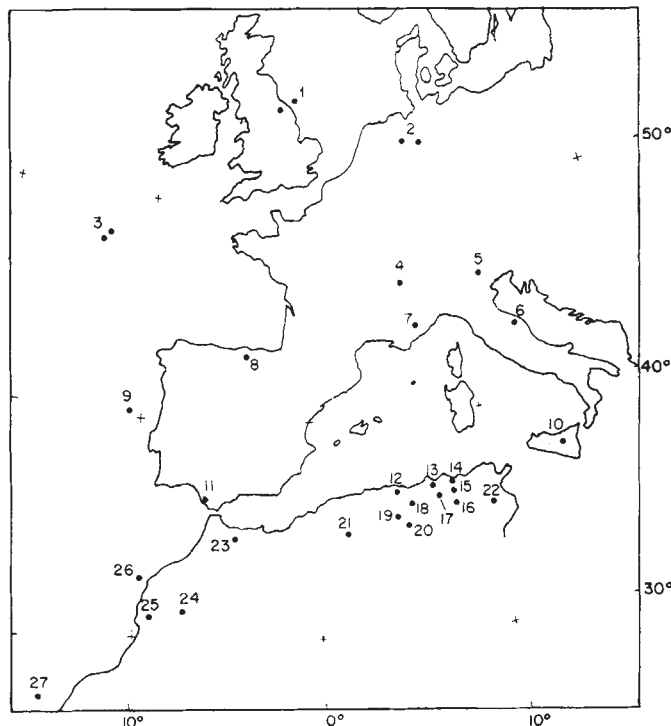


Fig. 4 Geographical position of various outcrops and boreholes with recorded Cenomanian–Turonian anoxic layers. (1) 'Black Band' of Yorkshire and Humberside, England and in the adjacent North Sea². Plenus Marls²¹. (2) Munster Basin and region of Hannover, West Germany⁵. (3) DSDP Sites 549 and 551³. (4) Flysch des Gets, Haute Savoie, France⁷. (5) Trento zone, Italy¹⁵. (6) Livello Bonarelli in Marche, Italy^{7,14}. (7) Flysch à Helminthoïdes, Alpes Maritimes, France and Italy⁷. (8) Cretaceous Navarro-Cantabric zone, northern Spain²⁰. (9) DSDP Site 398, Iberian margin. (10) Reitano Nappe, Argille Scagliose Varicolori, Sicily and Lucania, Italy⁸. (11) Andalusia, Sub-betic units, Spain⁷. (12) Massylian flysch, Petite Kabylie, Algeria^{9–12}. (13) Massylian flysch, Grande Kabylie, Algeria. (14) Massylian flysch, Constantine area, Algeria. (15) Edough Massif, Algeria^{9–12}. (16) Ain Fakroum and Ain Babouche^{9–12}. (17) Eastern Algeria^{9–12}. (18) Biban district, Algeria^{9–12}. (19) Western Biban–Hodna district^{9–12}. (20) Djebel Fernane, SW of Bou Saada, Algeria^{9–12}. (21) Eastern Ouarsenis^{9–12}. (22) Bahloul horizon, Central Tunisia¹⁵. (23) Chouamat and Melloussa flysch units⁸. (24), (25) The Cretaceous Atlas Gulf in Morocco⁶. (26) DSDP sites 370 and 416; here the Cenomanian–Turonian horizons have been swept off by subsequent erosion³. (27) DSDP site 369; Cenomanian carbonaceous layers are reworked here in layers of later Cretaceous age³.

level, perhaps 450 m higher than nowadays² of eustatic origin¹ which caused a large-scale transgression of the continents, for it is easily conceivable that the increased surface area offered to biogenic and terrigenous sediments during the maximum of the transgression, and the accompanying reduction of the area subjected to subaerial erosion, made less sediment available for sedimentation in each unit of time and surface area. Jenkyns⁷ has even suggested that the Cretaceous transgression was important in fostering marine anoxia, and coincidence of the CTBSH with the maximum extension of a worldwide transgression cannot be denied. But every major transgression, that of the Maestrichtian for example, have not induced ocean-wide anoxic events. During late Cenomanian–early Turonian times, special conditions must have prevailed in the oceans, with a general stagnation of sea-water below depths of a few hundred metres. The duration of this event may have been very short, of the order of a few hundred thousand years, as in the North European epicontinental seas or at the foot of the North American margin, but much longer elsewhere. High oceanic fertility seems to have favoured a conspicuous CTBSH and high TOC content. But high fertility was not essential for its development. Thus the CTBSH, probably unique in the Mesozoic–Cenozoic strati-

graphic record, resulted from the convergence of favourable conditions that cannot be incorporated in general and simple model and which cannot easily be identified in delicately balanced environments as these of the oceans.

Received 25 July; accepted 3 November 1983.

- Vail, P. R., Mitchum, R. M. & Thompson, S. *Am. Assoc. petrol. Geol. Mem.* **26**, 83–98 (1971).
- Hancock, J. M. & Kauffman, E. G. *J. geol. Soc.* **136**, 175–186 (1979).
- de Graciansky, P. C. *et al. Revue Institut Français du Pétrole*, **37**.
- de Graciansky, P. C. *et al. Init. Rep. DSDP 80* (in the press).
- Hart, M. B. & Bigg, P. J. *I.G.C.P. Project 58*, 177–185 (1981).
- von Rad, U., Hinz, K., Sarntheim, M. & Seibold, E. *Geology of the Northwest African Continental Margin* (Springer, Berlin, 1982).
- Jenkyns, H. C. *J. geol. Soc.* **137**, 171–188 (1980).
- Grandjacquet, C. & Mascle, G. in *The Ocean Basins and Margins* Vol. 4B, 257–329 (Plenum, New York, 1978).
- Caire A. *Publ. Serv. Carte Géologique Algérie* **16**, 1–818 (1957).
- Mattauer, M. *Publ. Serv. Carte Géologique Algérie* **17**, 1–534 (1958).
- Emberger, J. *Publ. Serv. Carte Géologique Algérie* **27**, 1–398 (1960).
- Hilly, J. *Publ. Serv. Carte Géologique Algérie* **19**, 1–408 (1962); **36**, 1–134 (1967).
- Burrollet, P. F., Mugniot, J. M. & Sweeney, P. in *The Ocean Basins and Margins* Vol. 4B, 331–359 (Plenum, New York, 1978).
- Wezel, F. C. *Ateneo Parmense, acta naturalia* **15**, 243–259 (1979).
- Bosellini, A., Broglio Loriga, C. & Busetto, C. *Riv. Ital. Paleont.* **84**, 897–946 (1978).
- Arthur, M. A. & Premoli Silva I. in *Nature and Origin of Cretaceous Carbon-Rich Facies* 7–54 (Academic, New York, 1982).
- Joides Journal*, vol. IX, no. 1, 1983, p. 22.
- Waples, D. W. *Init. Rep. DSDP 80* (in the press).
- Bouillin J.-P. *Mém. Soc. Géol. Fr.* **135**, 1–84 (1978).
- Lamolda, M. A. *Cahiers de Micropal.* **4**, 21–27 (1978).
- Institute of Geological Sciences and Norwegian Petroleum Directorate*, Rep. no. 77–25, 26 (1977).
- Herbin, J. P. & Deroo, G. *Bull. Soc. Géol. France, 7^{me} Serie* **24**, 497–510 (1982).
- Einsele, G. & Weidman, J. in *Geology of the Northwest African Continental Margin* (eds von Rad, U. *et al.*) 396–414 (Springer, Berlin, 1982).
- Stamm, R. & Thein, J. in *Geology of the Northwest African Continental Margin* (eds von Rad, U. *et al.*) 458–474 (Springer, Berlin, 1982).

Black smoker chimney fragments in Cyprus sulphide deposits

Elisabeth Oudin

Bureau de Recherches Géologiques et Minières, Avenue de Concy, Orléans-la Source (Loiret), B.P. 6009, 45060 Orléans, Cedex, France

Georges Constantinou

Cyprus Geological Survey, Nicosia, Cyprus

Hydrothermal vent fragments have been identified in the brecciated ore from the massive sulphide deposits of the Troodos complex, Cyprus. They display mineral associations, textures and zonations similar to those observed in recently discovered hydrothermal vents forming on the ocean floor near the East Pacific Rise^{1–3}, the Galapagos ridge^{4,5}, the Juan de Fuca ridge⁶ and the Guaymas basin (Gulf of California)⁷. The high temperature 'black smoker' chimney walls are especially well preserved in Cyprus ore. Traces of fauna similar to that found near present-day vents are also present in these Cyprus samples. The differences between the present-day and fossil vents are mainly due to ageing.

Present-day sulphide deposits are situated near the axis of medium- to fast-spreading submarine ridge sections at an average depth of 2,600 m. The vents which consist of porous aggregates of predominant sulphates and/or sulphides, form at the seawater–basalt interface when the high temperature (up to 350 °C) hydrothermal fluid⁸ comes in contact with the cold (+2 °C) seawater. The hydrothermal fluid pervading the heavily fissured oceanic crust forms poorly zoned massive lenses or mounds resting on the oceanic crust⁹. Thin-zoned chimney walls, a few millimetres to a few centimetres wide, build up around the high-temperature fluid escaping from larger channel ways near 21° N and 13° N on the East Pacific Rise (EPR) and in the Guaymas basin. The chimneys, 15 m high and 30 cm in diameter, are called black or white smokers according to the colour of the mineral precipitating in the fluid in contact with seawater: that is, predominantly pyrrhotite in higher temperature black smokers and baryte and opal in lower

temperature white smokers. General preliminary descriptions of samples collected near 21° N (refs 9–12) on the Galapagos ridge¹³ and on the Juan de Fuca ridge¹⁴ have already been published. Detailed mineralogical studies are also available for 21° N (refs 15 and 16) and Juan de Fuca samples¹⁷.

Cyprus sulphide deposits are situated in the upper part of the Troodos ophiolite sequence which is generally considered to be a fragment of obducted oceanic crust^{18,19}. The deposits are intercalated between the upper and lower pillow lavas which are of late Cretaceous age^{20,21}. The mineralization forms massive sulphide lenses in which gangue minerals are extremely scarce^{22–25}. Pyrite is predominant and frequently associated with chalcopyrite. Both chalcopyrite and zinc sulphides are locally abundant. The predominant species (Cu, Fe or Zn sulphides) also vary from one deposit to another in the EPR.

Both the EPR and Cyprus deposits are characterized by their similar mineral association, mineral habits and intergrowths, and the presence of rare minerals such as jordanite ($\text{Pb}_{14}\text{As}_6\text{S}_{26}$) or idaite^{9,26}. Because of ageing processes the unstable phases characteristic of present-day deposits are absent or rare in the Cyprus deposits, having been dissolved or replaced or having re-equilibrated. For example, the unstable Zn-bearing isocubanite intergrown with Zn-bearing iron-rich chalcopyrite in EPR samples¹⁵ is probably replaced by stoichiometric chalcopyrite containing sphalerite exsolutions in Cyprus¹⁶ agreeing well with experimental data²⁷. Pyrrhotite and wurtzite are also rare in Cyprus ore. Isocubanite (CuFe_2S_3) is a high-temperature cubic phase referred to in the literature as 'ISS', chalcopyrhotite, cubic cubanite or cubanite II. The name isocubanite has been submitted to the IMA commission.

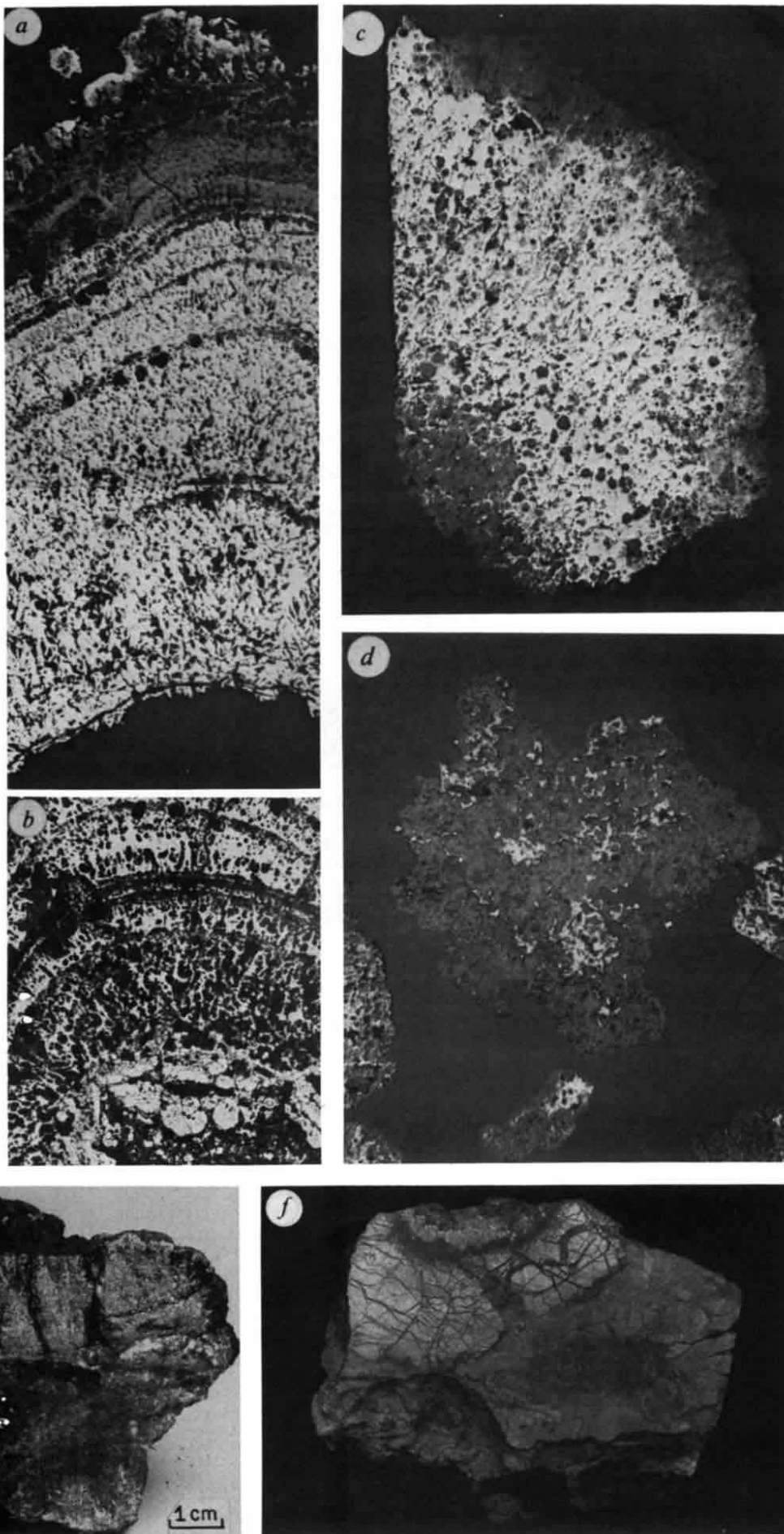
Typical vent fragments including high-temperature chimney ('black smokers') and 'mound' samples described in detail in EPR 21° N deposits^{15,28} or massive porous ore as observed in the Galapagos deposit¹³ have been identified in samples collected from various deposits in Cyprus (including Skourotis, Moussoulos, Mathiatis, Peristerka, Pitharokma and Agrokippia)^{9,29}. Best examples of vent fragments have been observed in Peristerka. EPR 21° N high temperature (250 °C–350 °C) chimney walls are zoned (Fig. 1a). An outer wall consisting predominantly of anhydrite is present on young active chimneys; it tends later to be replaced by sulphides and completely dissolve when the hydrothermal activity ceases. The inner wall is formed by an arborescent porous aggregate of euhedral sulphides. Several types of chimney wall can be distinguished in 21° N samples according to the predominant sulphide species. In the chalcopyrite-dominated type of chimney wall two types of zonation are observed. The primary zonation arises from variations in mineral grain size (Fig. 1a) and to more subtle variations in the mineral compositions such as the change in Fe content of chalcopyrite across the chimney wall¹⁵. The secondary zonation arises from early high-temperature oxidation of the active chimney. The seawater invades the porous wall and an oxidation sequence results: chalcopyrite (inner zone) is replaced by bornite, digenite (chalcocite) and covellite (outer zone); the excess Fe precipitates *in situ* as sulphides, oxides and hydroxides. The oxidation sequence is observed only in the outer zone of the sulphide wall in contact with the outer anhydrite wall. The thin inner chimney walls, few millimetres up to a maximum of 5 cm, are fragile and break down easily when the hydrothermal activity stops. The effect of the low-temperature submarine alteration process can be observed best in the samples from the inactive zones which have been lying on the sea floor for a longer period. In 21° N samples from the inactive zone, numerous dissolutions, replacements and neoformations (mainly sulphates) are observed¹⁶. A Cu–Fe sulphide (isocubanite/iron-rich chalcopyrite) dominated chimney wall is wholly or partly replaced by digenite and covellite (Fig. 1c and d). The whole surface as well as fissures in the chimney fragment in contact with seawater is altered. The fossil chimney wall fragments observed in Cyprus display characteristic features of equivalent EPR 21° N chimney walls such as porous arborescent aggregates of euhedral chalcopyrite grains²⁶ and a zonation due

to variation in mineral grain size (Fig. 1b). The thickness of Cyprus chimney wall fragments varies from a few millimetres to approximately 5 cm as in the 21° N deposits. As may be expected, the outer anhydrite wall (Fig. 1e and f) is absent and the primary zonation due to variation of chalcopyrite iron-content across the wall is not observed in Cyprus. Iron-rich chalcopyrite has been produced experimentally, but is very rarely observed in fossil ore deposits; its absence in Cyprus ore is, therefore, interpreted as reflecting re-equilibration during ageing (the stable stoichiometric chalcopyrite replacing the unstable iron-rich chalcopyrite). The Cyprus fragments display an alteration rim (Figs 1b, 1f and 2a) with a similar oxidation sequence to that observed in 21° N samples (Fig. 1a and 1c). A thin discontinuous rim of goethite sometimes replaces the sulphides or coats the fragments. Goethite is a frequent alteration product of EPR samples especially during the late low-temperature phase. However, the early high-temperature and the late low-temperature phases cannot be easily distinguished from one another since they produce the same mineral association. Therefore, the asymmetrical alteration rim observed in some Cyprus samples is tentatively interpreted as a combination of both types of alteration (Fig. 1f). Furthermore, a recent supergene selective alteration, in which the pyrite matrix is unchanged, cannot be completely ruled out because the Cyprus cupriferous deposits have been extensively mined since historical times and they lie very close to the land surface.

Other types of vent fragments are also found in Cyprus ores. Layered dendritic pyrite walls are present in both the Cyprus and 21° N deposits. Mound samples from 21° N, which are essentially formed of porous aggregates of zinc sulphides (sphalerite and/or wurtzite), are also present in Cyprus (Fig. 2a). The zinc sulphide spherulites may contain typical radiating inclusions of galena or jordanite as observed in 21° N and Juan de Fuca samples^{15,17}. Mound fragments are also slightly altered and the zinc sulphide is partly replaced by microcrystalline covellite. The alteration of the Cyprus vent fragments indicates that they lay on the sea floor prior to being embedded into a fine-grained matrix of pyrite and silica which tend to replace them (Fig. 1f). Pyrite associated with opal represents the late low-temperature phase in 21° N samples¹⁵. The late replacement of sulphides by silica has also been reported in the lower zone of some of the Cyprus ore bodies²⁵. Silica is isotropic under crossed Nicols in 21° N samples (opal A). In Cyprus samples the opal CT (ref. 30), chalcedony and quartz are interpreted as the result of recrystallization during ageing of the deposit. The fine-grained pyrite probably also represents a recrystallized collomorphic pyrite associated with opal A in EPR samples. Vent fragments are obvious in approximately 15% of over one hundred polished sections examined of Cyprus ore samples. However, it should be remembered that the samples were collected prior to the discovery of sulphides on the EPR and consequently the occurrence of vent fragments was not investigated *in situ*. The most common samples collected in Cyprus deposits are massive porous aggregates of either collomorphic or euhedral pyrite very similar to samples collected on the Galapagos ridge. Inclusions of pyrrhotite partially or completely replaced by pyrite, of wurtzite (or hexagonal zinc sulphide) are present in euhedral pyrite from both Galapagos and equivalent Cyprus samples. Chalcopyrite-dominated chimney wall fragments similar to 21° N are also included in Galapagos massive pyrite (Fig. 2b). The fragments are not oxidized and may have been rapidly included and protected by pyrite preventing submarine alteration. Chalcopyrite crystals lining pyrite cavities, where several generations of euhedral and collomorphic pyrite are observed, and sometimes associated with gypsum are also common to both Galapagos and Cyprus samples. Worm tubes and other organic remains which are abundant at the surface of some EPR vents where they can be replaced by opal and pyrite are also observed in Cyprus ores (Fig. 2c and d).

A previous mineralogical comparison²⁶ based on the mineral association and the mineral facies and intergrowths between EPR and Cyprus sulphide deposits is now extended by the

Fig. 1 *a, b*, Primary zonations in black smoker walls from EPR 21° N (*a*) and Peristerka (Cyprus) (*b*) due to variation in chalcopyrite grain size and outlined by dissolution zones (black). Photographed in polished section ($\times 3.5$). *c, d*, Black smoker fragments from 21° N inactive zone consisting of isocubanite and iron-rich chalcopyrite assemblage (white) at different stage of alteration in a fine-grained mixture of digenite and covellite (grey). Photographed in polished section (*c*, $\times 7$; *d*, $\times 6.5$). *e*, High-temperature (350 °C) active chimney wall (black smoker) in EPR 21° N deposit. The inner wall (light grey zone at the top of photograph) in contact with the hydrothermal fluid is formed of a porous aggregate of Cu-Fe sulphides (isocubanite and iron-rich chalcopyrite). The outer wall (dark grey zone) in contact with seawater is formed of anhydrite with minor sulphides. *f*, Black smoker fragment (lighter grey zone) in Peristerka ore which consists of a porous aggregate of euhedral chalcopyrite. The alteration rim (black) outlining the fragment is best developed on one side of the fragment (top) and may represent a combination of both early (high temperature) and late (low temperature) submarine alteration. The thinner alteration rim (bottom) may be due only to the low-temperature alteration stage. Silica and pyrite include and partially replace the fragment.



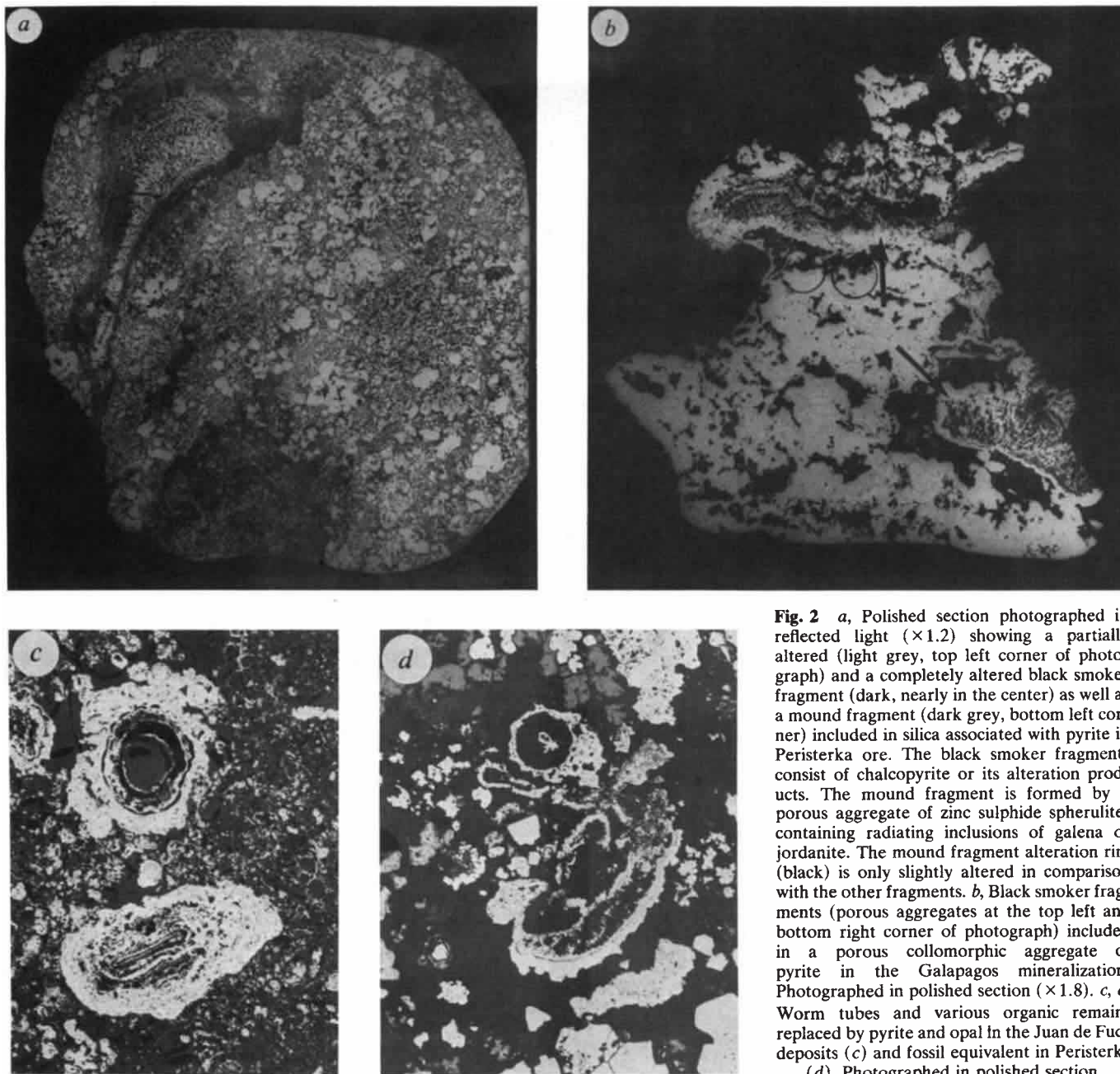


Fig. 2 *a*, Polished section photographed in reflected light ($\times 1.2$) showing a partially altered (light grey, top left corner of photograph) and a completely altered black smoker fragment (dark, nearly in the center) as well as a mound fragment (dark grey, bottom left corner) included in silica associated with pyrite in Peristerka ore. The black smoker fragments consist of chalcopyrite or its alteration products. The mound fragment is formed by a porous aggregate of zinc sulphide spherulites containing radiating inclusions of galena or jordanite. The mound fragment alteration rim (black) is only slightly altered in comparison with the other fragments. *b*, Black smoker fragments (porous aggregates at the top left and bottom right corner of photograph) included in a porous collomorphic aggregate of pyrite in the Galapagos mineralization. Photographed in polished section ($\times 1.8$). *c*, *d*, Worm tubes and various organic remains replaced by pyrite and opal in the Juan de Fuca deposits (*c*) and fossil equivalent in Peristerka (*d*). Photographed in polished section.

identification of vent fragments in Cyprus ore samples. Several types of vent fragments are present including the typical and spectacular black smoker (high-temperature) chimneys observed near 21°N on the EPR. Fossil worm tubes or other organic remains, abundant on some EPR vents, have also been found in Troodos. The massive porous aggregate of pyrite, very common in Cyprus ore, resembles closely the sample collected from the Galapagos ridge. The thin alteration rim present around the Cyprus vent fragments indicates that they lay on the sea floor. A complete alteration and subsequent early disappearance of the hydrothermal concretions was prevented by the late low temperature deposition of silica and pyrite. Although the vent fragments are relatively small compared to the original whole edifices, they are well preserved and the textures and zonations are easily recognized when directly compared with their present-day equivalents. The Cyprus deposits are unmetamorphosed and the post-deposition modifications can therefore be principally attributed to submarine alteration, diagenesis and ageing. Re-equilibration of unstable phases (isocubanite and iron-rich chalcopyrite) or recrystallization of amorphous silica are also interpreted as a result of ageing.

Comparisons of the mineralogy and texture of Cyprus ore with those of EPR deposits confirm that both were formed at the sea water-basalt interface as the result of hydrothermal circulations (in the same temperature range) in the oceanic crust near the ridge axis, in good agreement with geochemical and isotopic data³¹⁻³⁹.

We thank R. Hekinian, J. Edmond, H. Holland, J. Bishoff, W. Normark and A. Malahoff for samples from EPR deposits. Some of the Cyprus samples were collected by J. Bouladon and G. Pouit. Hand specimens on Fig. 1*e* and *f* were photographed by J. J. Poulin. We also thank J. Danto and M. Lanau for their helpful collaboration, H. Dunlop for helping with the translation and D. Verhelst for typing the manuscript.

Received 4 October 1983; accepted 3 January 1984.

1. Francheteau, J. *et al.* *Nature* **277**, 523-528 (1979).
2. Spiess, F. N. *et al.* *Science* **207**, 1421-1433 (1980).
3. Hekinian, R. *et al.* *EOS* **62**, 913 (1981).
4. Law, S., Malahoff, A., Embley, R., Fordari, D. *EOS* **62**, 1027 (1981).
5. Malahoff, A. *Mar. Technol. Soc. J.* **16**, 39-45 (1982).
6. Normark, W., Morton, J. L., Koski, R. A., Clague, D. A. & Delaney, J. R. *Geology* **11**, 158-163 (1983).
7. Lonsdale, P. F., Bishoff, J. L., Burns, U. M., Kastner, M. & Sweeney, R. E. *Earth planet. Sci. Lett.* **49**, 8-20 (1980).

8. MacDonald, K. C., Becker, K., Spless, F. N. & Ballard, R. D. *Earth planet. Sci. Lett.* **48**, 1-7 (1980).
9. Oudin, E. *Chron. Mines*, No. 470, 43-55 (1983).
10. Picot, P. *et al. Docum. Bur. Rech. Geol. Min.* No. 20 (1980).
11. Haymon, R. & Kastner, M. *Earth planet. Sci. Lett.* **54**, 363-381 (1981).
12. Styr, M. M. *et al. Earth planet. Sci. Lett.* **53**, 382-390 (1981).
13. Malahoff, A., Embley, R. W., Cronan, D. S. & Skirrow, R. *Mar. Min.* **4**, 123-137 (1983).
14. Normark, W. *et al. Mar. Technol. Soc. J.* **16**, 46-53 (1982).
15. Oudin, E. *Docum. Bur. Rech. Geol. Min.*, No. 25 (1981).
16. Oudin, E. *Mar. Min.* **4**, 39-72 (1983).
17. Koski, R. A., Clague, D. A. & Oudin, E. *Bull. Geol. Soc. Am.* (in press).
18. Gass, I. G. & Smewing, J. D. *Nature* **247**, 26-29 (1973).
19. Gass, I. G. & Smewing, J. D. *Sea* **7**, 339-362 (1981).
20. Mantis, M. *Cyprus Res. Cent.* **111**, 227-241 (1970).
21. Vine, F. J., Poster, C. K. & Gass, I. G. *Nature phys. Sci.* **244**, 34-38 (1973).
22. Constantinou, G. & Govett, G. J. S. *Trans. Instn Min. Metal.* **81**, B34-B46 (1972).
23. Constantinou, G. *Econ. Geol.* **68**, 83-86 (1973).
24. Constantinou, G. *Geol. Ass. Can. spec. Pap.* **14**, 187-209 (1976).
25. Constantinou, G. *Proc. Int. Ophiolite Symp.* (Min. Agn. Nat. Res. Geol. Surv. dept. Nicosia, Cyprus 1980).
26. Oudin, E., Picot, P. & Pouit, G. *Nature* **291**, 404-407 (1981).
27. Hutchinson, M. N. & Scott, S. D. *Econ. Geol.* **1**, 143-153 (1981).
28. Oudin, E. *et al.* (in preparation).
29. Oudin, E. & Constantinou, G. *Terra Cognita* **3**, 169-170 (1983).
30. Jones, J. B. & Segnit, E. R. *J. geol. Soc. Aust.* **18**, 57-68 (1971).
31. Spooner, E. T. C. & Fyfe, W. S. *Contr. Mineral. Petrol.* **42**, 287-304 (1973).
32. Spooner, E. T. C., Beckinsdale, R. D., Fyfe, W. S. & Smewing, J. D. *Contr. Mineral. Petrol.* **47**, 41-62 (1974).
33. Heaton, J. H. E. & Sheppard, S. M. F. *Volcanic Processes in Ore Genesis* 42-57 (Instn. Min. Metall., London, 1977).
34. Sheppard, S. M. F. *Volcanic Processes in Ore Genesis* 25-41 (Instn. Min. Metall., London, 1977).
35. Spooner, E. T. C. & Bray, C. T. *Nature*, **266**, 808-812 (1977).
36. Spooner, E. T. C., Chapman, H. J. & Smewing, J. D. *Geochim. cosmochim. Acta* **41**, 873-890 (1977).
37. Spooner, E. T. C. *Volcanic Processes in Ore Genesis* 68-71 (Institution of Mineralogists and Metallurgists, London, 1977).
38. Chapman, H. J. & Spooner, E. T. C. *Earth planet. Sci. Lett.* **35**, 71-78 (1978).
39. Spooner, E. T. C. *Geol. Ass. Can. spec. Pap.* **20**, 685-704 (1980).

Taphonomy, fauna, environment and ecology of Upper Sivaliks (Plio-Pleistocene) near Chandigarh, India

Rajan Gaur

Department of Anthropology, Panjab University,
Chandigarh 160014, India

S. R. K. Chopra

Panjabi University, Patiala, India

The fossil fauna of the molassic Sivalik deposits is well documented but little is known of its environment and ecology. Initial reports¹⁻⁸ were based on published fauna and flora, and palaeoecologically oriented field studies⁹⁻¹³, particularly of the Upper Sivalik Subgroup, were few. We present here a preliminary report on the fauna, flora, taphonomy, environment and ecology of Upper Sivaliks of north-east Chandigarh, India. Previously these have yielded cercopithecoids, *Procynocephalus*¹⁴ and *Theropithecus*¹⁵, and Palaeolithic artefacts^{16,17}. We found that the Upper Sivaliks are basically fluvatile. The Tatrot and Pinjors were laid down by laterally shifting rivers on a flood plain. The Boulder Conglomerates are predominantly alluvial fan deposits. The Upper Sivalik bone assemblage is highly sorted with the dominance of heavier, denser and more durable parts. The fauna and flora indicate a mosaic dominated by wooded grassland with bushland and grassland for Tatrots and dominant grassland with some bushland and wooded grassland for Pinjors. The geology and fauna indicate that the climate was warm and humid during Tatrots, relatively arid and less warm during Pinjors, and became increasingly cold and arid during Boulder Conglomerate deposition in this area.

A survey of the area (Fig. 1) by R.G. for a total period of 8 months from 1976 to 1980 yielded a large number of fossil invertebrates and vertebrates (Table 1). All the three formations of Upper Sivaliks—Tatrot, Pinjor and Boulder Conglomerate—are exposed here. Tatrots occur as discontinuous isolated patches separated by Pinjor outcrops. Clay, silt, sandstone and conglomeratic beds of various textures and hues, with cross-bedding, convolute bedding, load casts, desiccation fissures, silt

Table 1 Fossil vertebrates collected from the Upper Sivaliks near Chandigarh

Faunal type	Upper Sivaliks		
	Tatrot	Pinjor	Boulder Conglomerate
Pisces			
Seluridae indet.	+	+	—
Reptilia			
Crocodilidae indet.	+	+	—
Chelonidae indet.	+	+	—
Mammalia			
Rodentia			
Muridae indet.	0	1(1)	0
Carnivora			
Vishnuictis sp. nov.	0	1(1)	0
Proboscidea			
Stegodon bombifrons	1(1)	1(1)	0
Stegodon insignis	3(1)	14(3)	0
Elephas planifrons	1(1)	20(4)	0
Elephas platycephalus	0	2(1)	0
Elephas hysudricus	0	12(3)	0
Perissodactyla			
Equidae			
Equus sivalensis	0	66(7)	0
Equus subsp. nov.	0	5(1)	0
Rhinocerotidae			
Coelodonta platyrhinus	0	6(2)	0
Rhinoceros sivalensis	0	9(3)	0
Rhinoceros palaeindicus	0	7(2)	0
Artiodactyla			
Suidae			
Potamochoerus theobaldi	0	1(1)	0
Potamochoerus sp.	1(1)	0	0
Sus falconeri	0	8(2)	0
Sus sp. nov.	0	2(1)	0
Hippopotamidae			
Haxaprotodon sivalensis	11(3)	5(1)	0
Cervidae			
Rucervus simplicidens	0	2(1)	0
Cervidae indet.	0	12(2)	0
Camelidae			
Camelus sivalensis	0	7(1)	0
Giraffidae			
Sivatherium giganteum	2(1)	8(1)	0
Bovidae			
Leptobos falconeri	4(1)	0	0
Demelops palaeindicus	0	3(1)	0
Bovini indet.	16(2)	95(10)	0
Reduncini indet.	0	4(1)	0
Alcelaphini indet.	0	14(3)	0
Antilopini indet.	0	14(2)	0
Total	40(11)	317(56)	0

+ Present; — absent. Values in parentheses represent the minimum number of individuals.

lenses, scour and fill structures, mudballs, fining-upwards sequences characterize the Upper Sivaliks of this area. A preliminary magnetostratigraphic study¹⁸ suggests an age of 3.15–0.73 Myr for the Upper Sivaliks of this area. However, equivalent rocks in northern Pakistan have been assigned an age of 5.5–0.6 Myr¹⁹. Both these studies^{18,19} place the boundary between Tatrot and Pinjor beds at the Gauss/Matuyama transition, that is, about 2.5 Myr.

The Upper Sivaliks in this area display three major types of facies; flood-plain, channel and alluvial fan. The mean (M_z) of the grain size of Tatrot sandstones is less than that of Pinjor sandstones (2.6 and 2.3 units respectively). The finer character of Tatrot deposits suggests that the energy level of the depositing water was low compared with the Pinjor and Boulder Conglomerate formations, probably due to lower gradients. Large-scale flat bedding in Tatrots suggests deposition on extensive flood-plains. Fining-upwards sequences were observed in the Tatrot and Pinjor deposits indicating deposition by the laterally shifting channels of meandering rivers. Red coloration of sediments generally implies warm and humid conditions²⁰. Decreasing red coloration of sediments from Tatrot to Boulder Conglomerate Formation probably denotes a decrease in warmth and humidity from Tatrots to Pinjors and relatively cold and arid to semi-arid climate for the Boulder Conglomerates. The change from a warm to a relatively cool climate from Tatrots to Pinjors is confirmed by the decrease in the carbonate content of the rocks

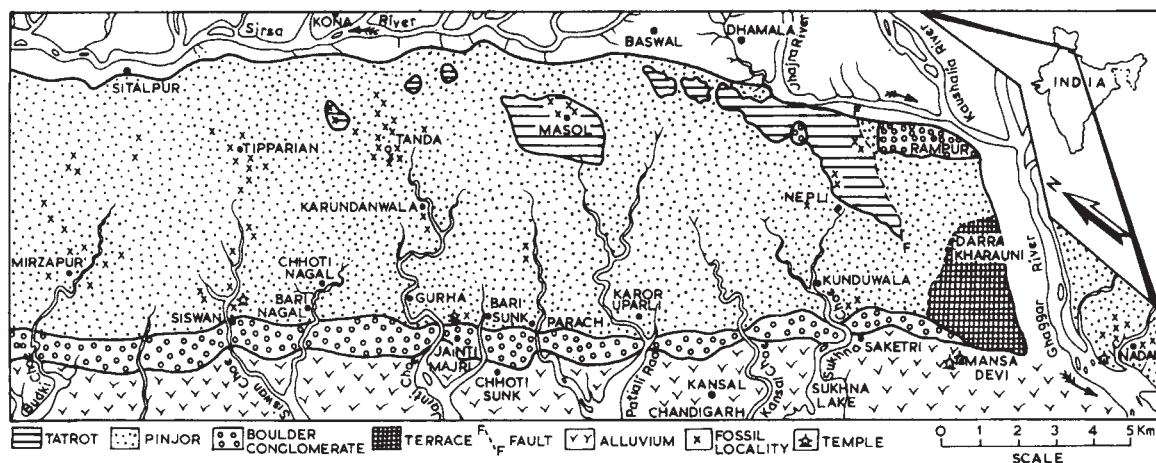


Fig. 1 Generalized locality map of the Upper Sivaliks.

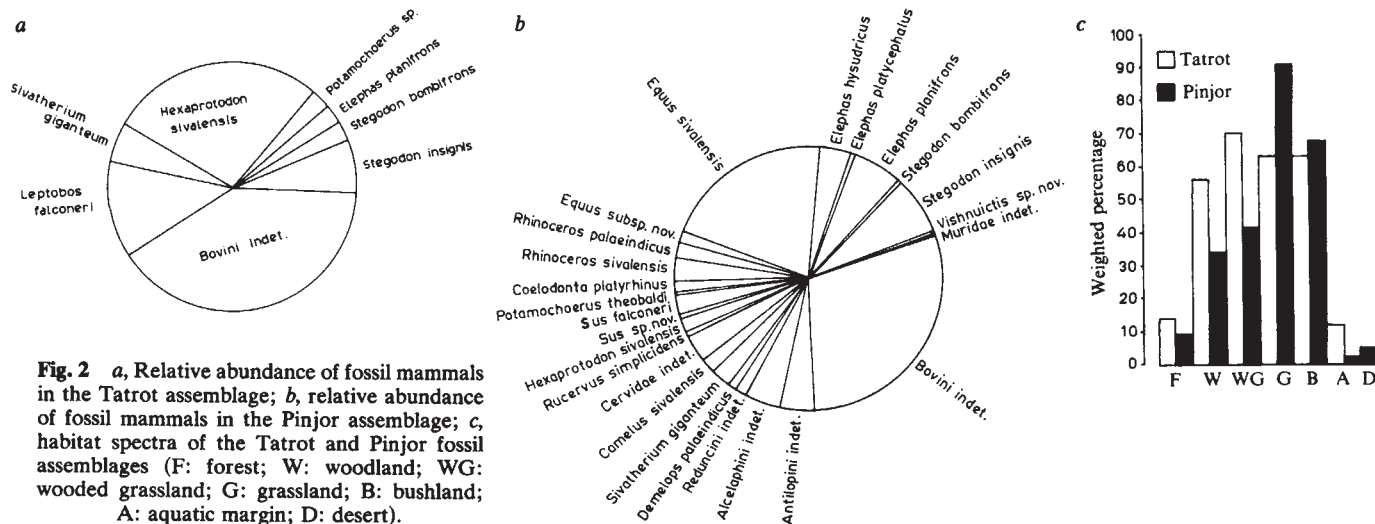
from 7.3% to 2.7%, since deposition of carbonates generally takes place in warmer climates: at low temperatures water contains more carbonate in dissolved form²⁰. The lithology of Pinjors suggests that they were deposited in channels, on the flood-plain and also in ox-bow lakes, ponds and other depressions of the flood-plain. From the base to the top of Pinjors there is a gradual increase in the coarseness of deposits suggesting a steady elevation in the source area. The increase in coarseness of sediments from the Tatrot to Boulder Conglomerate could indicate a steady change in climate from more humid during Tatrot to relatively arid during Pinjors, becoming increasingly arid towards the top of Pinjors. A sudden and sharp elevation of the adjacent mountains probably marked the beginning of the Boulder Conglomerate Formation. Consequently, thick alluvial fan aprons, assisted by a subsiding basin, accumulated at the foot of the mountains. The sinuosity of the depositing riverine system decreased and a nearly braided system developed. Since the best-developed alluvial fans now occur in arid to semi-arid and subarctic regions²¹, the climate during the deposition of the Boulder Conglomerates was probably arid to semi-arid and relatively cold.

Most of the skeletal elements of the fossil assemblages found are disarticulated and fragmentary: the bones do not occur in proportions that would be expected if whole animals were preserved. From the proportional representation of the skeletal elements (number present/number expected in an average skeleton, based on the minimum number of individuals) (Table 2), it is clear that the heavier, denser and more durable elements, such as, mandibles, metapodia and isolated teeth, have been preferably preserved. The poor representation of vertebrae, ribs and phalanges indicates that certain taphonomic processes

operated against the preservation of relatively light and fragile bones. On the whole, the proportional representation of skeletal elements is similar to the second and third dispersal groups of Voorhies²², denoting that the bone assemblage was sorted.

Pinjor sediments have yielded more fossils than the Tatrots. Identifiable fossils are rare in the Boulder Conglomerate Formation. We found that the Tatrot/Pinjur boundary (which corresponds with the Gauss/Matuyama transition) in this area is marked by the first appearance of *Equus sivalensis*, *Equus* subsp. nov., *Elephas hysudricus*, *Sus* sp. nov., *Vishnuictis* sp. nov., *Rhinoceros palaeindicus*, *Rhinoceros sivalensis* and *Rucervus simplicidens*. *Hipparion*, which has been recorded from the Pinjur Formation of northern Pakistan^{19,23}, is absent in the Pinjors of this area. In contrast to Opdyke *et al.*¹⁹ and Barry *et al.*²³, we found *Elephas planifrons* in the Tatrot beds near Masol belonging to the Gauss magnetic epoch¹⁸. As in the Upper Sivaliks of Pakistan¹⁹, cervids with antlers make their first appearance in the Pinjors of this area also. In the present assemblage, *Camelus* is restricted to the Pinjur Formation, though it has earlier been recorded from Tatrot Formation also^{24,25}. Three new taxa (to be described elsewhere), *Vishnuictis* sp. nov., *Sus* sp. nov. and *Equus* subsp. nov., are reported from the Pinjur Formation of this area. *Potamochoerus theobaldi* is reported from this area for the first time and *Stegodon bombifrons* is recorded for the first time from Pinjors.

In addition to vertebrates, a number of freshwater gastropods, crustaceans, ostracods (*Candona*, *Ilyocypris*, *Zonocypris*, *Potamocypris* and *Hemicypris*) and unionids were collected. The floral remains found are restricted to a few unidentified seeds and numerous oogonia of charophytes, that is, *Sphaerochara*



sp., *Chara rantzieni*, *Chara surazpurica*, *Hornichara maslovi* and *Nitellopsis* sp.

The abundance of cold-blooded animals, such as crocodiles and turtles, in Tatrot suggests a well-watered landscape and a warm climate. The occurrence of charophytes in some Tatrot beds indicates clear water bodies with little current action and high pH. The abundance of *Hexaprotodon* in the Tatrot assemblage (Fig. 2a) suggests large sluggish rivers and a relatively warm climate. There are few typical cursorial and grassland animals in the Tatrot assemblage. The Tatrot bovids show varied adaptations ranging from grassland to bushland, savannas and woodlands. Tatrot proboscideans and suids indicate the presence of bushlands and wooded grasslands. There are no species indicative of forest. The faunal diversity spectra of Tatrot mammals (Fig. 2c) suggests a mosaic dominated by wooded grassland with grassland, bushland and some woodland.

The chelonids and crocodilians are less abundant in Pinjors than Tatrots. Since the reptiles attain great size and profusion in warmer climates²⁰, their low frequency in Pinjors probably indicates a colder climate than that of Tatrots. The occurrence of charophytes, freshwater ostracods, gastropods and unionids indicates the presence of localized water bodies, such as ponds and ox-bow lakes. The most abundant Pinjor mammals are equids, bovids and elephants (Fig. 2b)—all grazing animals—suggesting that a major part of Pinjor habitat was probably comprised of grassland. Dominance by animals with cursorial adaptations, equids and bovids, suggests an open landscape. The habitat spectra of Pinjor mammals (Fig. 2c) also supports the view that grassland dominated the habitat which also included some bushland and wooded grassland. Forest and closed habitat forms are rare. There is a change from a dominant wooded grassland/bushland habitat of Tatrots to an open grassland in Pinjors probably resulting from increasing aridity and also indicated by the desert mammals in the Pinjors (Fig. 2c).

Since identifiable fossils are rare in the Boulder Conglomerate Formation we cannot discuss its palaeoecological conditions.

The Plio-Pleistocene Upper Sivalik molassic deposits were deposited by a well-developed meandering riverine system on a broad flood-plain. Thick aprons of alluvial fans were deposited during Boulder Conglomerate times and the climate became more arid and cold. The taphonomic processes during Upper Sivaliks favoured the preservation of heavier, denser and more durable elements resulting in a highly sorted bone assemblage. The geology and fauna suggest a warm and humid climate for Tatrots and a relatively arid and less warm climate for Pinjors.

We thank Dr Peter Andrews, British Museum of Natural History, London, for his comments on an earlier version of this paper.

Table 2 Proportional abundance of various skeletal elements in the Upper Sivalik fossil assemblage in north-east Chandigarh

Skeletal element	Number	Proportional percentage
Maxillae	3	2.24
Mandibles	18	13.43
Maxillary	100	9.32
Mandibular } Teeth	125	9.33
Skull	1	1.50
Vertebrae	12	0.64
Ribs	7	0.87
Humerii	4	2.90
Radius/ulnae	2	1.40
Femora	6	4.40
Tibiae	8	5.97
Astragali	9	6.71
Calcanea	5	3.73
Podials (other than astragalus and calcaneum)	16	1.08
Metapodia (3rd metapodial of perissodactyls and cannon bone of artiodactyls)		
Metacarpals	11	13.40
Metatarsals	14	17.07
Phalanges (including metapodia of suids and hippos)	10	0.35

Received 25 October 1983; accepted 12 January 1984.

1. Krynine, P. D. *Am. J. Sci.* **34**, 422 (1937).
2. Tattersall, I. *Nature* **221**, 451 (1969).
3. Tattersall, I. *Nature* **224**, 821 (1969).
4. Prasad, K. N. *Nature* **232**, 413 (1971).
5. Prasad, K. N. in *World Anthropology—Paleoanthropology, Morphology and Palaeoecology* (Mouton, The Hague, 1975).
6. Gaur, R., Vasishat, R. N. & Chopra, S. R. K. *Recent Res. Geol.* **5**, 206 (1978).
7. Gaur, R., Sabhlok, A. K., Chopra, S. R. K. & Suneja, I. J. *Man Envir.* **2**, 69 (1978).
8. Vasishat, R. N., Gaur, R. & Chopra, S. R. K. *Palaeogeogr. Palaeoclim. Palaeoecol.* **23**, 131 (1978).
9. Halstead, L. B. & Nanda, A. C. *Bull. Indian geol. Ass.* **6**, 63 (1973).
10. Badgley, C. & Behrensmeier, A. K. *Palaeogeogr. Palaeoclim. Palaeoecol.* **30**, 133 (1980).
11. Behrensmeier, A. K. & Tauxe, L. *Sedimentology* **29**, 331 (1982).
12. Gaur, R. & Chopra, S. R. K. *Man Envir.* **7**, 39 (1983).
13. Gaur, R. & Chopra, S. R. K. *Palaeogeogr. Palaeoclim. Palaeoecol.* **43**, 313 (1983).
14. Verma, B. C. *J. palaeont. Soc. India* **13**, 53 (1969).
15. Gupta, V. J. & Sahni, A. *Bull. Indian geol. Ass.* **14**, 69 (1981).
16. Sharma, J. C. *Curr. Anthropol.* **18**, 94 (1977).
17. Verma, B. C. *J. geol. Soc. India* **16**, 518 (1975).
18. Azzaroli, A. & Napoleone, G. *Riv. ital. Paleont. Stratigr.* **87**, 739 (1982).
19. Opdyke, N. et al. *Palaeogeogr. Palaeoclim. Palaeoecol.* **27**, 1 (1979).
20. Schwarzbach, M. *Climates of the Past* (Van Nostrand, London, 1963).
21. Reineck, H. E. & Singh, I. B. *Depositional Sedimentary Environments* (Springer, New York, 1973).
22. Voorhies, M. *Contr. Geol. Univ. Wyoming* **1**, 1 (1969).
23. Barry, J. C., Lindsay, E. H. & Jacobs, L. L. *Palaeogeogr. Palaeoclim. Palaeoecol.* **37**, 95 (1982).
24. Sahni, M. R. & Kahn, E. J. *Palaeont. Soc. India* **4**, 59 (1964).
25. Nanda, A. C. *J. geol. Soc. India* **19**, 160 (1978).

New chronological data on European Plio-Pleistocene faunas and hominid occupation sites

N. Thouveny & E. Bonifay

Laboratoire de Géologie du Quaternaire, CNRS, Case 907 Faculté des Sciences de Luminy 13288, Marseille Cedex 9, France

The relative chronology of the continental European Plio-Pleistocene deposits has generally been deduced from the biostratigraphic data with support from radiometric ages derived by K/Ar or fission track methods. In many cases it is impossible to obtain reliable radiometric dates, and in such cases palaeomagnetic studies can be a useful tool for the correlation of the local faunas into the magnetic polarity time scale. We present here a study of the remanent magnetization of lacustrine sediments and basaltic lavas of the Velay region (Massif Central, France) associated with early and late Villafranchian faunas and prehistoric remains.

Volcanic activity in the Massif Central (France) has been paroxysmic from the Miocene almost to the present (ceasing some 3,450 yr ago. However, in some parts of that area, for example Western Velay, it began only about 3.5 Myr ago and ended about 0.7 Myr ago¹. Basaltic and basaltic flows were laid down on eroded surfaces of the Palaeozoic belts and on the Oligocene lacustrine clays. Barred valleys and Maar's craters were occupied by lakes during the Pliocene and Pleistocene. The lacustrine sediments containing fossils were overlaid and protected from the erosion by other lava flows.

In the region of Le Puy en Velay, palaeontological discoveries began during the last century². Thus the Viallette, Sainzelles and Soleihac deposits have been well known for more than 100 years. Figure 1 shows the geographical situation of the sites. Many stratigraphic studies have been carried out on the sites³⁻⁸. Maar structures were definitely recognized at Sainzelles and Soleihac but at Viallette this origin is not fully confirmed. At each site, the study of mammal bones allows us to recognize the fossil assemblages, to assess their level of evolution and compare them with other well-known fossil faunas. The Viallette fauna is thought to be the oldest Villafranchian fauna in Europe⁹. The Sainzelles fauna is considered one of the last Villafranchian faunas¹⁰ or one of the first post-Villafranchian faunas⁹. The sorting and fracturing of the bones led us to think of it as one of the oldest sites of hominid occupation in Europe^{11,12}. The Soleihac fauna represents the transitional term between late



Fig. 1 Geographical localization of sites.

Villafranchian faunas and middle Pleistocene faunas. The fossils lay around the oldest known built structure made by the European hominids, a rocky wall made with basalt blocks more than 20 m in length. The stone tool industry—producing choppers, chopping-tools, proto-biface implements (one is known), and scrapers is certainly older than the Abbeville culture¹². A study is under way to try to correlate the magnetic polarities of those formations with the standard geomagnetic polarity time scale¹³ using all the reliable chronological results, including K/Ar, fission tracks and palaeontology.

The stratigraphy of each site studied is summarized in Fig. 2. The Pliocene and Pleistocene deposits lie in general on the Oligocene clays. In Viallette (Fig. 2a), the fossiliferous sediments lie on a basaltic flow and were probably overlain by the upper flow (now partially eroded). K/Ar dating gave ages of 3.30 ± 0.06 Myr for the lower flow and 2.60 ± 0.1 Myr for the upper flow⁷. The fission track age (on sphene crystals contained in the sediments) was 3.14 ± 0.6 Myr (ref. 14). A preliminary palaeomagnetic study of 12 samples of sediments gave a stable normal polarity thought to have been acquired during the Gauss epoch.

In Sainzelles (Fig. 2b), the fossiliferous deposits lie on a pyroclastic flag (tuff-ring) and are overlain by pyroclastics and a basaltic flow (E.B., J. P. Raynal, J. Mergoill and N.T., unpublished data). The sediments are, from bottom to top: colluviums of pyroclastic origin interlain with sands; pink lacustrine clays; silts and clays with fossil mammals. K/Ar dating of the upper flow gave two ages: 1.3 Myr and 1.4 Myr (refs 15, 16).

In Soleihac (Fig. 2c), the lacustrine deposits are contained in a maar depression of about 1.0 km diameter. The deposits are, from bottom to top: blocks and gravels of the tuff-ring (bed F, 20 m thick); blue clays interlain with colluviums of pyroclastite (bed E, 50–80 m thick); silts and sands of fluviolacustrine origin (bed D, 10–15 m thick); grey silts and sands containing the palaeontological and prehistoric remains (bed C, 1–2 m thick); sands and gravels alternating with organic layers (bed B, 2–3 m thick); basaltic fallen rocks and colluviums (bed A, 10–25 m thick). This stratigraphy will be completed by the study of two long cores (60 to 80 m) drilled near the prehistoric site; the first results show an extension of the top of the deposits above bed B.

About 250 samples orientated with a magnetic compass were analysed at the Laboratoire de Géologie du Quaternaire du Centre National de la Recherche Scientifique in Marseille-Luminy. The measurements were made with a Digico spinner magnetometer¹⁷. Analysis of the natural remanent magnetization (NRM) was made using several methods: alternating field (a.f.) demagnetization (with a Schonstedt GDS-1 demagnetizer), 300 °C thermal demagnetization (30 min in a Digico furnace), viscosity test¹⁸, use of the demagnetization vector diagrams^{19,20} and remagnetization circles²¹. All palaeomagnetic conclusions are the result of a whole decomposition of the NRM in order to test its stability.

The mean directions of characteristic magnetizations and the statistical parameters for the various sites are presented in Table 1. We discuss here the final conclusions of analysis at each site.

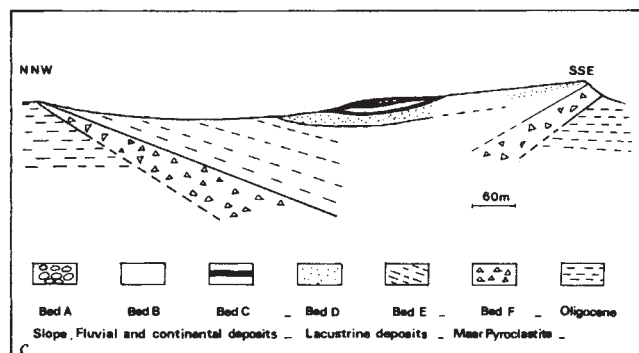
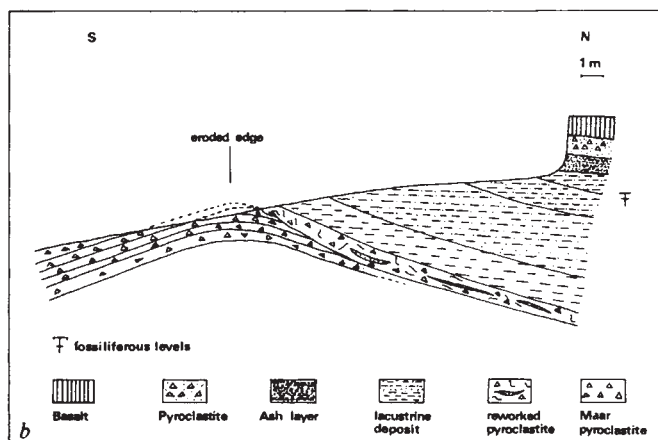
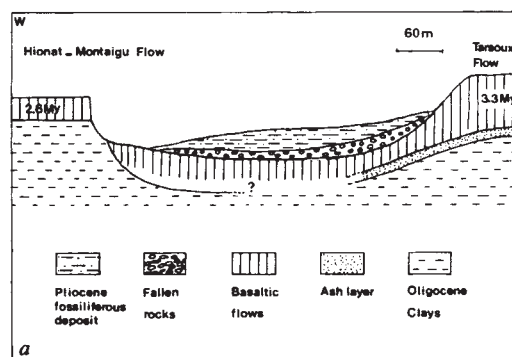


Fig. 2 a, Stratigraphical scheme of the Viallette's deposits. b, Stratigraphical scheme of the Sainzelles's deposits. c, Stratigraphical scheme of the Soleihac's deposits.

Viallette: The polarity of the lower flow was clearly reversed after 60 mT (millitesla) a.f. treatment. The sediments showed a stable normal polarity after 30 mT a.f. treatment and 300 °C thermal treatment, which can be attributed to the characteristic magnetization. The polarity of the upper flow was clearly reversed after 60 mT a.f. treatment but the mean direction is significantly different from that of the lower flow. Clearly the sediments were laid down in a normal geomagnetic field; the lavas were laid down in a reverse geomagnetic field.

The K/Ar age of the lower flow (3.33 ± 0.06 Myr) would correspond to the normal interval at the beginning of the Gauss epoch but its reverse polarity does not agree with that age. However, we suggest two reverse intervals near 3.3 Myr: one at the end of the Gilbert Epoch (3.80–3.40 Myr) or the Mammoth Event in the Gauss Epoch (3.15–3.05 Myr). The fission track age of the sediments (3.14 ± 0.60 Myr) goes from the end of Gilbert to the end of Gauss. The normal polarity restricts the choice to the Gauss Epoch (3.40–2.48 Myr) with the exception of the Mammoth and Kaena Events. The K/Ar age of the upper flow (2.60 ± 0.1 Myr) would correspond to the normal interval at the end of the Gauss Epoch. The reversed polarity does not agree with that age, but the following intervals are possible: in the Gauss Epoch, Mammoth Event (3.15–3.05 Myr)

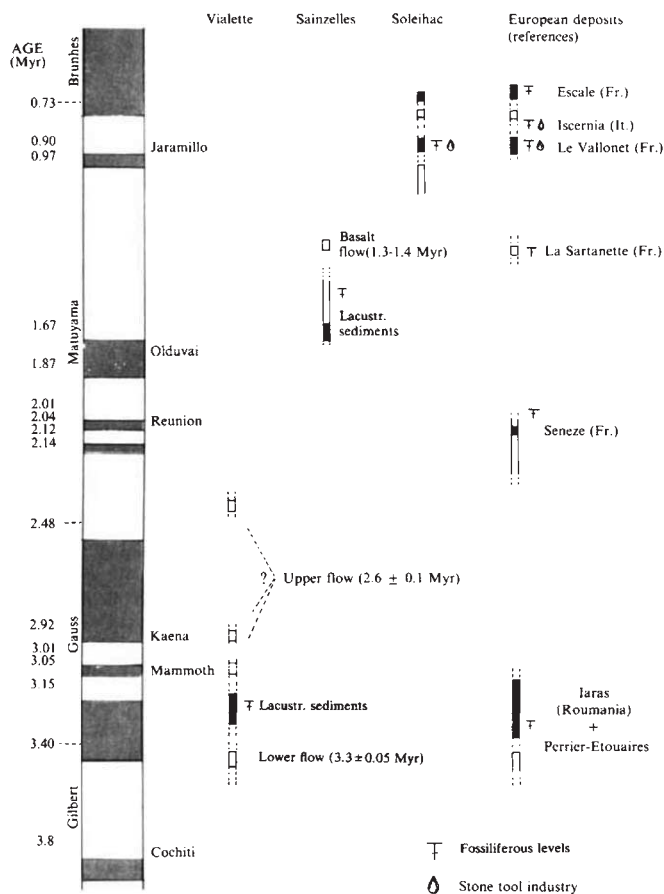


Fig. 3 Correlation of the fossiliferous sequences at Vialette, Sainzelles and Soleihac with the geomagnetic polarity time scale¹³. Black areas are normal polarity intervals, white areas are reverse polarity intervals.

or Kaena Event (3.01–2.92 Myr); beginning of the Matuyama Epoch (2.48–2.14 Myr). The third proposition seems to be in best agreement with the K/Ar age.

The biostratigraphy can help us to determine in which part of Gauss Epoch the fossiliferous sediments can be placed. A faunal association very similar to the Vialette one has been discovered in Romania in lacustrine deposits of the Brasov depression near Iaras²². Palaeomagnetic study of two sections allowed the lower section of reverse polarity to be placed in the Gilbert Epoch and the upper section of normal polarity to be placed in the Gauss Epoch; a short reverse event was recorded at the top part of the upper section attributed to Mammoth or Kaena. As the fauna lies at the bottom of the upper section its age is thought to be between 3.2 and 3.3 Myr. Comparison with the well-known early Villafranchian fauna of Perrier-Etoudaires (the twin site of Vialette) gives no help as its age is very controversial: either 2.50–2.60 Myr or 3.30–3.40 Myr (refs 23–25). Reference to the Iaras deposits provides a new argument in favour of the oldest age for the first appearance of the Villafranchian faunas in Europe: 3.3 Myr. Thus, we will retain the interval 3.4–3.15 Myr, beginning of the Gauss Epoch, for the Vialette fossiliferous sediments.

Sainzelles: Two polarities have been discovered in the lacustrine deposits: the lower levels (pink clays) showed a normal stable remanent magnetization after 30 and 50 mT a.f. treatment and 300 °C thermal treatment, which is considered as the characteristic depositional magnetization. The upper levels (top of pink clays and above) showed an intermediate direction after 30 mT treatment which is probably the result of a reverse primary depositional magnetization covered by a hard viscous component (total demagnetization of samples showed no tendency towards reverse directions). This hypothesis is strongly

Table 1 Characteristic values of the mean directions after demagnetization 30 mT a.f. or 300 °C for sediments and 60 mT a.f. for lava flows

	Mean direction		Statistical parameters		No. of samples
	<i>D</i>	<i>I</i>	<i>K</i>	Alpha 95	
<i>Vialette</i>					
Lower flow	187	−65	27	7	16
Upper flow	227	−87	33	7.6	12
Sediments	354	52	10.7	7	42
<i>Sainzelles</i>					
Sediments (lower levels)	1	63	37.2	5.5	19
Flow	175	−55	6,100	1.1	4
<i>Soleihac</i>					
(top of bed E and bed D)	8	53	16.3	4.4	63

The repartition of the directions for Sainzelles sediment upper levels and Soleihac sediment lower levels (bottom of bed E) cannot be considered as fisherian (K lower than 3). D , declination; I , inclination; K , precision parameter of Fisher; Alpha 95, radius of the cone of confidence at 95%.

supported by the clearly reversed polarity of the thermoremanent magnetization of the overlying flow. The K/Ar ages obtained on the flow show that the normal polarity of the lower deposits has to be attributed to the Olduvai Event (1.87–1.67 Myr). The reversed polarity must have been acquired afterwards (between 1.67 and 1.40 or 1.30 Myr).

The Sainzelles fauna belongs to the last interval. It shows many similarities to the fauna of the La Sartanette (France) cave, located near Nîmes²⁶. This fauna is contained in lacustrine deposits of reverse polarity²⁷, lying on a conglomerate derived from a terrace of the late Pliocene or early Pleistocene. This agrees well with the interval agreed for the Sainzelles fauna which was previously defined as intermediate between the fauna of Seneze and Le Vallonet^{9–10}. The maximum age of the Seneze fauna is 2.01 Myr (refs 28, 29) and the age of Le Vallonet fauna is between 1.20 and 0.80 Myr (ref. 30), or more precisely, it corresponds to the Jaramillo Event³¹. The K/Ar age of the Sainzelles flow restricts the upper limit to 1.40 or 1.30 Myr and the palaeomagnetic data restrict the lower limit to 1.67 Myr.

Soleihac: The lower part of bed E (bottom of the lacustrine deposits) showed a great scatter between the individual directions of magnetization after a.f. treatment. The thermal treatment gave reverse or intermediate directions. This suggests that a reverse primary magnetization was covered by a hard normal viscous component. The top of bed E and bed D showed a very stable normal magnetization which is considered as a characteristic depositional one. At the top of the long cores (above bed B) two samples showed a stable reverse magnetization at 5.2 m depth, two showed an intermediate magnetization at 5 m depth, and two showed a stable normal magnetization at 3 m depth. Thus, the Soleihac deposits have preserved, from bottom to top, a succession of four polarity zones: reverse–normal–reverse–normal. The fauna and the prehistoric remains lie in the first normal polarity zone.

The Soleihac fauna can be compared with the fauna of another well-known European site, Isernia in Italy: the two associations contain common species. The Isernia fauna also contains younger species, known in another French site, the l'Escalade cave where the deposits are younger than the Brunhes–Matuyama boundary (0.73 Myr)³². The Isernia fauna has a degree of evolution between those of Soleihac and l'Escalade and its age is recognized as older than the Brunhes–Matuyama boundary^{33,34}. Furthermore, if the Soleihac fauna is considered more evolved than that of Vallonet, it is not older than 0.97 Myr. Therefore, the normal interval corresponding to the faunal and prehistoric levels in Soleihac must be regarded as the Jaramillo Event, a well-known marker of the end of the Villafranchian over Europe and India³⁵.

The results of this magnetostratigraphic study, according to radiometric ages and biostratigraphic levels, are shown in Fig. 3. The Viallette fauna is thus attributed to the beginning of the Gauss Epoch (3.40–3.15 Myr) which agrees with previous estimates. The Sainzelles fauna with hominid evidence is younger than the end of the Olduvai Event (1.67 Myr) but older than about 1.30 Myr. In Soleihac, the fauna and the prehistoric levels correspond to the Jaramillo Event (0.97–0.90 Myr). This agrees well with the age of the deposits of Le Vallonet and Isernia. This period corresponds to a well known critical moment for the Pleistocene environment but also for hominid technology as evidenced by the oldest 'architectural' work in Europe.

This work was supported by the Ministère de la Recherche, contract DGRST no. 81 D 0378 to E.B.

Received 8 September; accepted 20 December 1983.

1. Bout, P. & De Goer, A. in *La Préhistoire française* Vol. 1 (ed. de Lumley, H.) 304–310 (CNRS, Paris, 1976).
2. Boule, M. *Bull. Serv. Carte géol. France* 4, 28–259 (1892).
3. Bout, P. *Geol. Mijnbouw* 26, 83–93 (1964).
4. Bout, P. *Palaeogeogr. Palaeoclimat., Palaeoecol.* 8, 95–106 (1970).
5. Bonifay, E. & Tiercelin, J. J. *C. r. hebdom. Séanc. Acad. Sci., Paris D284*, 2455–2457 (1977).
6. Bonifay, E. & Bonifay, M. F. in *Le Bassin du Puy aux Temps préhistoriques*, 19–36 (Musée Crozatier, Le-Puy-en-Velay, 1981).
7. Bandet, Y., Donville, J. & Michaux, J. *Bull. Soc. Géol. Fr.* 20, 245–251 (1978).
8. Couthures, J. *Bull. Ass. Fr. Et. Quat.* 4, 171–173 (1979).
9. Heintz, E., Guerin, C., Martin, R. & Prat, F. *Mém. B.R.G.M.* 78, 169–182 (1974).
10. Bonifay, M. F. in *Le Villafranchien méditerranéen*, 105–115 (Université Lille, 1982).
11. Bonifay, M. F. in *Le Quaternaire*, 41–45 (CNRS/INQUA, Paris, 1973).
12. Bonifay, E. *La Recherche* 128, 1442–1444 (1981).
13. Mankinen, E. A. & Dalrymple, G. B. *J. geophys. Res.* 84, 615–625 (1979).
14. Biquand, D., Cassignol, C., Chambaudet, A. & Couthures, J. *Bull. Ass. Fr. Et. Quat.* 2, 83–87 (1981).
15. Azzaroli, A. *G. Geol., Bologna* 35, 1–21 (1967).
16. Bellon, M. thesis, Univ. Orsay (1971).
17. Molyneux, L. *Geophys. J.* 24, 429–433 (1971).
18. Thellier, E. & O. *Ann. Geophys.* 1, 37–52 (1944).
19. Zijdeveld, J. D. A. in *Methods in Palaeomagnetism* (eds Collinson, D. W., Creer, K. M. & Runcorn, S. K.) 254–286 (Elsevier, Amsterdam, 1967).
20. Dunlop, D. Y. *Phys. Earth. planet. Int.* 20, 12–24 (1970).
21. Halls, H. C. *Phys. Earth. planet. Int.* 16, 1–11 (1978).
22. Ghenea, C. *et al.* in *SEQS-INQUA Congr.*, 31–33 (Inst. Geol. Geophys., Bucarest, 1981).
23. Semah, F. & Biquand, D. *Bull. Ass. Fr. Et. Quat.* 1, 83–96 (1979).
24. Couthures, J. & Pastre, J. F. in *Le Villafranchien méditerranéen*, 179–185 (Université Lille, 1982).
25. Ly, M. H., Cantagrel, J. M., De Goër, A. & Vincent, P. M. in *Le Villafranchien méditerranéen*, 497–424 (Université Lille, 1982).
26. Bonnet, A. & Malaval, M. *C. r. hebdom. Séanc. Acad. Sci. Paris D282*, 2155–2157 (1976).
27. Thouveny, N., Taieb, M. & Bonnet, A. *Boreas* (submitted).
28. Prevot, M. thesis, Univ. Paris (1971).
29. Opdyke, N. D., Lindsay, E. & Kukla, G. *Nature* 304, 365 (1983).
30. Michelot, C. in *Le Villafranchien méditerranéen*, 123 (Université Lille, 1982).
31. Renault-Miskowsky, J. & Girard, M. in *La Préhistoire française* (ed. de Lumley, H.) 1, 469 (CNRS, Paris, 1976).
32. Bonifay, E., Bonifay, M. F. & Thouveny, N. in *Abstr. II^e Congr. INQUA, Moscow*, 28.
33. Coltorti, M., Cremaschi, M., Delitala, M. C. & Esu, D. *Nature* 300, 173–176 (1982).
34. Isernia la Pineta, un accampamento più antico di 700 000 anni (Calderini, Bologna, 1983).
35. Azzaroli, A. & Napoléone, G. *Riv. Ital. Paleont.* 87, 739–762 (1981).

Maternal dominance, breeding success and birth sex ratios in red deer

T. H. Clutton-Brock, S. D. Albon & F. E. Guinness

Large Animal Research Group, Department of Zoology, University of Cambridge, Cambridge CB2 3EJ, UK

That female vertebrates who are able to invest more than average in their offspring should produce male-biased sex ratios was first suggested by Trivers and Willard¹ who reasoned that the breeding success of males is more variable than that of females and so may be more strongly influenced by parental investment. Their evidence for a positive association between maternal condition and birth sex ratios was unconvincing^{2,3} and no subsequent studies have produced conclusive evidence supporting their hypothesis^{4,8}. We show here that, in polygynous red deer (*Cervus elaphus*), dominant mothers produce significantly more sons than subordinates and that maternal rank has a greater effect on the breeding success of males than females.

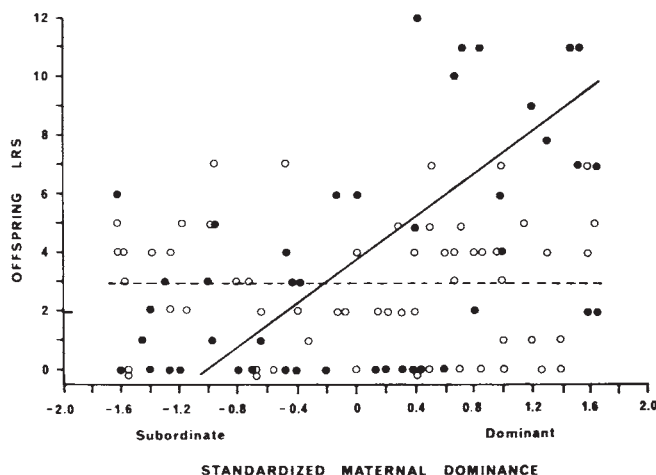


Fig. 1 Lifetime reproductive success (LRS) of sons and daughters of red deer hinds in relation to their mother's social rank. See text for details (●, males; ○, females). The slope shown is the reduced major axis³².

Records for around 120 red deer hinds and 100 stags on the Isle of Rhum (Scotland) collected between 1971 and 1983 show the breeding success of each individual and the proportion of male and female calves born to each mother⁹. The social dominance of each hind is measured by calculating the total number of other hinds more than 1 yr old that she has been seen to threaten or displace, weighted by their ranks, and divided by the total number of hinds that threaten or displace her, again weighted by their ranks^{9,10}. Previous analyses have shown that this index is closely related to other measures of dominance including the ratio of the number of older animals a hind dominates to the number of younger ones that dominate her¹⁰. Since the hind population was divided into four segments by preferred habitat, and hinds rarely interact with members of segments other than their own¹¹, ranks are calculated within each segment of the population, and then combined on a standardized scale. Estimates of rank calculated in this way are stable from year to year and are not affected by whether or not the hind has a calf at foot⁹.

Among hinds, most components of breeding success are related to dominance rank, except for fecundity (the proportion of years in which a hind produces a calf; Table 1). Dominant hinds also conceive and change coat earlier than subordinates. As the timing of both coat change and conception is known to be influenced by body condition in red deer^{12,13}, this indicates that, as in primates¹⁴, dominant animals are of a superior body condition compared with subordinates. Dominance rank is also correlated with the hind's body size ($r = 0.397$, $t_{24} = 2.116$, $P < 0.05$) as well as with the social rank of her mother ($r = 0.201$, $t_{102} = 2.069$, $P < 0.05$).

Previous analysis of sex ratio data from Rhum failed to show that birth sex ratios varied significantly with the mother's age, parity or home-range area^{9,15}. However, comparison of the sex ratios produced by hinds allocated to three equal groups on the basis of rank, revealed a significant relationship between a hind's dominance and the sex ratio of her offspring (Table 2a: $G_2 = 6.586$, $P < 0.05$). The sex ratio of offspring born to high-ranking females is significantly male-biased (60.6% male, $n = 149$; binomial test; $z = 3.25$, $P < 0.01$) while the sex ratio of offspring born to medium- and low-ranking hinds does not deviate significantly from parity (medium-ranking mothers: 53.9% male, $z = 0.929$, not significant; low-ranking mothers: 46.9% male, $z = 0.868$, not significant). This difference is consistent within all four segments in the study population and also within all seven cohorts of hinds born between 1966 and 1972 (Wilcoxon matched-pairs signed ranks test: $T = 0$, $P < 0.02$), showing that the relationship between rank and the sex ratio is not a consequence of any association between a hind's dominance and her age or the location of her home range.

Finally, we examined the correlation between the rank of hinds and the birth sex ratio of their progeny, using a sample of 50 mothers (see below). Since the number of calves per hind was small (3–12), we weighted each point by the number of calves on which it was based. This analysis, too, showed a significant association between a hind's rank and the sex ratio of her progeny ($r=0.291$, $t_{48}=2.106$, $P<0.05$).

If the relationship between maternal rank and the sex ratio is an adaptive one, a mother's rank should affect the breeding success of her sons more than that of her daughters. Our records indicate that this is the case. Measures of breeding success spanning at least half the average reproductive life are available for 58 female and 44 male offspring born between 1971 and 1974 to 50 different mothers of known dominance rank. Analysis of 16 males and 29 females born between 1967 and 1970

shows that individual differences in breeding success are consistent across the lifespan¹⁶ and that lifetime reproductive success (LRS) can be accurately predicted from a knowledge of breeding success over the first half of the average breeding lifespan (correlation and least-squares regression for LRS on breeding success up to 8 yr old for 16 stags: $r=0.937$, $t_{14}=9.994$, $P<0.001$; $y=1.07+0.17x$; correlation and least-squares regression for LRS on breeding success up to 8 yr old for 29 hinds: $r=0.873$, $t_{27}=9.301$, $P<0.001$; $y=0.46+1.65x$). Using these equations, we estimated the residual reproductive success of the animals in our sample that were still alive in 1983 (11 of the 44 stags and 37 of the 58 hinds) and added this to their observed breeding success during the 9 or more years for which we have followed them. (In practice, this involved only a small error, since the average remaining life expectancy was only 1.1 yr for the 11 stags and 2.3 yr for the 37 hinds.)

Figure 1 shows that breeding success of males was positively correlated with their mothers' ranks ($r=0.519$). Slopes were calculated using least-squares regression and reduced major axis since error was to be expected on both axes (least-squares regression: $t_{42}=3.936$, $P<0.001$; reduced major axis: $t=7.583$, $P<0.001$). In contrast, there is no consistent relationship between maternal rank and the breeding success of females ($r=0.027$, $n=58$; not significant). When the two relationships were compared, the slope for sons was significantly steeper than that for daughters (least-squares regression: $F_{1,98}=12.22$, $P<0.001$; reduced major axis, $z=6.450$, $P<0.001$). The breeding success of males born to mothers above median rank was greater than that of females (means: males, 5.14; females, 3.03; $t_{52}=2.12$, $P<0.05$), while females born to mothers of below median rank tended to be more successful than males (means: males, 2.00; females, 3.08; $t_{46}=1.69$, $0.1>P>0.05$).

Although the breeding success of sons increases with their mothers' rank, a substantial proportion of variance is unaccounted for by this analysis. In part, this is a consequence of calf mortality, which occurs throughout the range of maternal rank and accounts for the majority of zero scores. Rutting injuries also contribute to variance in male success. Six stags were permanently injured before they completed their reproductive lives and five of these were less successful than expected. Six others were born to mothers that failed to conceive in the year following their birth and who continued to suckle their offspring throughout their first year of life—unlike mothers who conceived again, who weaned their sons at between 6 and 8 months⁹. All six stags whose mothers failed to conceive in the year following their birth were more successful than was predicted on the basis of their mother's rank (binomial test, $P<0.032$).

Table 1 Measures of breeding performance in hinds above median rank versus those below it

	Subordinate		Dominant		
Median age of first breeding (yr)	3.90	(60)	3.48	(52)	$G=4.32$, $P<0.05$
Fecundity rate (yr ⁻¹)	0.65	(46)	0.68	(43)	$G=0.01$, NS
% Conceiving before median conception date	38	(55)	58	(50)	$G=4.91$, $P<0.05$
% Changing into winter coat before 1 November	35	(37)	62	(37)	$G=5.48$, $P<0.02$
Median weight of calves at birth (kg)	6.4	(55)	6.8	(45)	$G=14.38$, $P<0.001$
% Calf mortality during first year of life	45	(55)	28	(50)	$G=3.45$, $0.1>P>0.05$
Mean lifespan (yr)	9	(14)	11	(14)	Fisher's exact probability, $P<0.05$
Median lifetime reproductive success (calves reared to 1 yr old)	2.3	(14)	6.0	(15)	Fisher's exact probability, $P<0.05$

Data from cohorts born between 1967 and 1978. Sample sizes of mothers (shown in parentheses) differ between measures since not all variables were available for all individuals. The analyses control for the effects of age as well as for geographical variation in reproductive performance within the study area. All P values are for two-tailed t -tests.

Table 2 Sex ratio of calves in relation to mother's rank

a	Hind dominance rank			b	Upper (11)		Kilmory (37)		Intermediate (28)		Samhnsan Insir (22)	
	Low (34)	Medium (29)	High (35)		Sub.	Dom.	Sub.	Dom.	Sub.	Dom.	Sub.	Dom.
Males	61	90	149	Males	10	24	35	80	26	56	30	39
Females	69	77	97	Females	13	11	34	60	33	37	26	29
% Males	46.9	53.9	60.6	% Males	43.4	68.6	50.7	57.1	44.1	60.2	53.6	57.4
c	Year of hind's birth											
	1966 (6)		1967 (8)		1968 (8)		1969 (5)		1970 (8)		1971 (10)	
	Sub.	Dom.	Sub.	Dom.	Sub.	Dom.	Sub.	Dom.	Sub.	Dom.	Sub.	Dom.
Males	10	21	14	19	12	13	13	19	7	10	13	17
Females	10	11	13	12	14	10	15	9	6	2	10	11
% Males	50.0	65.6	51.9	61.3	46.2	56.5	46.4	67.9	53.8	83.3	56.5	60.7

a, All births 1970–82 from mothers born between 1957 and 1974; b, analysed separately for each of the four geographical subdivisions of our study population¹⁰; and c, analysed separately for hinds belonging to each of the cohorts born between 1966 and 1972. Numbers of hinds in each category are shown in parentheses. In the first analysis (a) hinds were divided into three equal divisions on the basis of rank. The minor inequalities in sample size were a consequence of tied values of rank. In b and c, sample sizes within categories were small and hinds were divided into two categories: above median rank for their segment of the population (dominant, dom.) and below it (subordinate, sub.).

The superior body condition and larger body size of dominant mothers may explain why their sons are more successful than those of subordinates: in red deer, the breeding success of stags depends on their fighting ability and body size⁹ which, in turn, are related to growth and nutrition during the first 18 months of life^{9,16,17}. Maternal rank may have a lesser effect on the reproductive success of daughters both because early growth and adult size are less important to females¹⁶ and because differences in nutrition during the first year of life have a weaker influence on growth in females⁹.

The mechanism underlying the association between rank and the sex ratio is unknown. In domestic ruminants, attempts to manipulate the sex ratio have been unsuccessful¹⁸, although significant trends have been previously reported in white-tailed deer¹⁹⁻²¹. In red deer, the absence of any difference in calving rate between dominant and subordinate hinds once breeding has begun (Table 1) suggests that variation in the sex ratio is unlikely to be due to differences in zygote mortality unless these occur sufficiently soon after conception for the hind to conceive again in the same breeding season.

In contrast to these results, recent studies of baboons and macaques have shown that dominant mothers produce significantly female-biased sex ratios while subordinates tend to produce more sons than daughters⁶⁻⁸. It is suggested that because female primates typically remain in their natal troops and inherit their mother's rank whereas males disperse after adolescence, maternal rank may exert a stronger influence on the breeding success of daughters than on that of sons⁸. Unfortunately, no quantitative evidence of the comparative effects of maternal rank on the breeding success of male and female offspring in primates has yet been published^{4,5}. However, the available evidence indicates that, in primates, body size may not be as important a determinant of male breeding success as in ungulates²²⁻²⁶, suggesting that early growth and parental investment may also be less important to male primates. Conversely, female dominance hierarchies are more clearly defined among primates and a female's rank is more strongly influenced by the status of her mother than in red deer²⁷⁻³¹; it would not be surprising, therefore, if in primates a mother's rank exerted a stronger influence on the breeding success of her daughters than in red deer.

We thank the director of the Nature Conservancy Council for permission to work on the island of Rhum; Jon Seger, Michael Reiss, Paul Harvey, John Maynard Smith, Michael Bulmer, Rick Charnov, Nick Davies, Anthony Arak, Doug Meikle and Dave Harper for comments and discussion; and Callan Duck and Glenn Jason, our collaborators in the Rhum red deer project, for their part in collecting data. The research is funded by grants from the SERC, NERC, Leverhulme Trust and the Royal Society.

Received 30 September 1983; accepted 4 January 1984.

1. Trivers, R. L. & Willard, D. E. *Science* **179**, 90-92 (1973).
2. Myers, J. H. *Am. Nat.* **112**, 381-388 (1978).
3. Williams, G. C. *Proc. R. Soc. B205*, 567-580 (1979).
4. Clutton-Brock, T. H. *Nature* **298**, 11-13 (1982).
5. Clutton-Brock, T. H. & Albon, S. D. in *Current Problems in Sociobiology* (ed. King's College Sociobiology Group) (Cambridge University Press, 1982).
6. Altmann, J. *Baboon Mothers and Infants* (Harvard University Press, Cambridge, 1980).
7. Simpson, M. J. & Simpson, A. E. *Nature* **300**, 440-441 (1982).
8. Silk, J. B. *Am. Nat.* **121**, 56-66 (1983).
9. Clutton-Brock, T. H., Guinness, F. E. & Albon, S. D. *Red Deer: Behavior and Ecology of Two Sexes* (University of Chicago Press, 1982).
10. Clutton-Brock, T. H., Albon, S. D., Gibson, R. M. & Guinness, F. E. *Anim. Behav.* **27**, 211-225 (1979).
11. Clutton-Brock, T. H., Albon, S. D. & Guinness, F. E. *Nature* **300**, 178-180 (1982).
12. Ryder, M. L. *J. Zool.* **181**, 137-143 (1977).
13. Mitchell, B. & Lincoln, G. A. *J. Zool.* **171**, 141-152 (1973).
14. Small, M. F. *Int. J. Primatol.* **2**, 91-95 (1981).
15. Clutton-Brock, T. H., Albon, S. D. & Guinness, F. E. *Nature* **289**, 487-487 (1981).
16. Clutton-Brock, T. H. in *Evolution from Molecules to Men* (ed. Bendall, D. S.) (Cambridge University Press, 1983).
17. Suttie, J. M. *Anim. Behav.* **31**, 610-611 (1983).
18. Charnov, E. L. *The Theory of Sex Allocation* (Princeton University Press, 1982).
19. Verme, L. J. *J. Wildl. Mgmt* **33**, 881-887 (1969).
20. Verme, L. J. & Ozoga, J. J. *J. Wildl. Mgmt* **45**, 710-715 (1981).
21. McCullough, D. R. *The George Reserve Deer Herd* (University of Michigan Press, 1979).
22. Geist, V. *Mountain Sheep* (University of Chicago Press, 1971).
23. Bygott, D. thesis, Univ. Cambridge (1974).

24. Novikoshi, K. & Koyama, N. in *Symp. of the 5th Congr. of Int. Primatological Soc.* (eds Konde, S., Kawai, M., Ehara, A. & Kawamura, S.) 43-61 (Japan Science, Tokyo, 1974).
25. Packer, C. thesis, Univ. Sussex (1977).
26. Dewsbury, D. A. *Q. Rev. Biol.* **57**, 135-159 (1982).
27. Kawai, M. *Primates* **14**, 113-159 (1958).
28. Missakiah, E. *Primates* **13**, 169-180 (1972).
29. Hausfater, G., Altmann, J. & Altmann, A. *Science* **217**, 752-755 (1982).
30. Dunbar, R. I. M. *Behav. Ecol. Sociobiol.* **7**, 253-265 (1980).
31. Clutton-Brock, T. H., Greenwood, P. & Powell, R. P. Z. *Tierpsychol.* **41**, 202-216 (1976).
32. Kermack, K. A. & Haldane, J. B. S. *Biometrika* **37**, 30-41 (1950).

Hyperpolarization of fish retinal horizontal cells by kainate and quisqualate

M. W. Hankins & K. H. Ruddock

Gunnar Svaetichin Laboratory of Neurophysiology,
Department of Pure and Applied Biology, Imperial College, Ascot,
Berkshire SL5 7PY, UK

Kainic (KA) and quisqualic (QA) acids have a potent depolarizing action on a variety of neurones of the central nervous system¹⁻⁴, including retinal horizontal cells^{5,6}. We now report the novel finding that at low concentrations (1-3 μ M), these 'excitatory' amino acids hyperpolarize horizontal cells of the fish retina. We show that the hyperpolarizing effects of both KA and QA are reversed by the γ -aminobutyric acid (GABA) antagonist bicuculline^{7,8}, whereas a second GABA antagonist, picrotoxin⁹, reverses the effects of KA, but not of QA. Neither GABA antagonist influences horizontal cell depolarization by 50 μ M KA or 50 μ M QA, thus the excitatory (depolarizing) and inhibitory (hyperpolarizing) effects of the amino acids involve independent mechanisms. We provide evidence that the hyperpolarizing effects are not mediated by the dopaminergic pathways associated with retinal horizontal cells¹⁰.

Retinal horizontal cells in cyprinid fish have large cell bodies and are organized into layers of interconnected neurones¹¹. In the dark, their resting potential is -20 mV to -40 mV and, in the main, they hyperpolarize in response to light, although C-type cells are depolarized by some stimulus wavelengths^{12,13}. It is proposed that in darkness, photoreceptors release a neurotransmitter which depolarizes the horizontal cells, and that light reduces the rate of transmitter release¹⁴, causing hyperpolarization of the horizontal cells. Depolarization of C-type cells is attributed to a GABAergic feedback connection between horizontal cells and photoreceptors^{15,16}. Horizontal cells are depolarized by ≥ 10 μ M concentrations of KA and QA, but are hyperpolarized by a third excitatory amino acid, *N*-methyl-D-aspartate, and in each case, the light-evoked responses (S-potentials) are suppressed¹⁷. We now report that at low concentrations, KA and QA hyperpolarize horizontal cells, apparently through a GABAergic pathway.

Retinas dissected from the excised eyes of dark-adapted cyprinid fish (roach, *Rutilus rutilus*) were placed receptor side up in a perfusion chamber and firmly held by a clamping ring and coarse-mesh net¹⁷. The individual retina was superfused at a rate of 1-4 ml min⁻¹, producing rapid changes in response to drugs. Microelectrodes (d.c. tip resistance of 80-150 M Ω) were introduced from above the retina, which was stimulated from below with 300-ms light flashes. All data refer to L-type units which hyperpolarize for all stimulus wavelengths. Figure 1a illustrates hyperpolarization of the horizontal cells and reduction of their S-potentials by 1 μ M and 3 μ M KA. The average change in membrane potential induced by 3 μ M KA was -10.0 ± 1.2 mV (± 1 s.e.), with the S-potential amplitude changed by a factor of 0.43 ± 0.07 (31 cells from 17 retinas). The corresponding values for 3 μ M QA are -9.7 ± 1.0 mV and 0.42 ± 0.06 (13 cells from 11 retinas). The figures exclude values for the first two or three applications of the amino acids, which were usually ineffective, and those for some 10% of the retinas in which

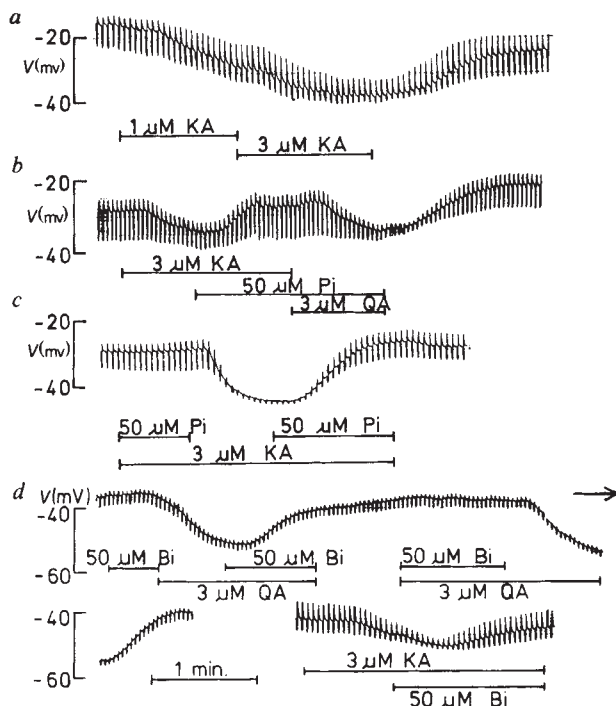


Fig. 1 Membrane potential (V) of L-type horizontal cells with time. *a*, Hyperpolarization of horizontal cells by $1\ \mu\text{M}$ and $3\ \mu\text{M}$ KA. *b*, *c*, Reversal of hyperpolarizing effects of KA, but not those of QA, by $50\ \mu\text{M}$ picrotoxin. *d*, Reversal of hyperpolarizing effects of $3\ \mu\text{M}$ KA and QA by bicuculline.

Methods: Hyperpolarizing responses (S-potentials) were elicited by weak ($0.04\ \mu\text{W mm}^{-2}$), 300-ms light flashes (wavelength 650 nm), repeated at ~ 3 -s intervals. Drugs were added to the standard Ringer (110 mM NaCl, 2.5 mM KCl, 20 mM NaHCO_3 ; 20 mM glucose, pH 7.7) as indicated by the bars below the traces. Bi, bicuculline; KA, kainic acid; Pi, picrotoxin; QA, quisqualic acid. The solutions containing Pi and Bi were magnetically stirred, and in the case of Bi, sonicated, to overcome low solubility. Note that the two records of *d* refer to different cells.

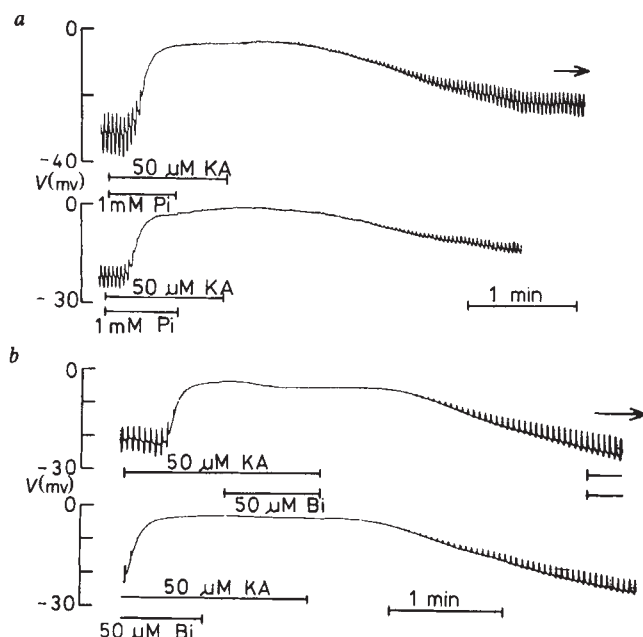


Fig. 2 Neither picrotoxin nor bicuculline reverses the depolarization of horizontal cells by KA. The arrows denote continuity between successive traces. Same protocol as for Fig. 1.

further applications also had no effect. The effects of $3\ \mu\text{M}$ KA, but not those of $3\ \mu\text{M}$ QA, were reversed by $50\ \mu\text{M}$ picrotoxin, a GABA antagonist⁹ (Fig. 1*b, c*), and picrotoxin at $600\ \mu\text{M}$ remained ineffective against $3\ \mu\text{M}$ QA. A second GABA antagonist, bicuculline^{7,8}, reversed the actions of both $3\ \mu\text{M}$ KA and QA (Fig. 1*d*), but neither picrotoxin nor bicuculline reversed the depolarization of horizontal cells by $50\ \mu\text{M}$ KA (Fig. 2) or QA. The picrotoxin concentration in Fig. 2*a* is $1\ \text{mM}$, giving the same KA/picrotoxin concentration ratio as in Fig. 1*b, c*, thus the antagonistic effects of picrotoxin are specific to the hyperpolarizing action of KA.

GABA depolarizes horizontal cells of the turtle¹⁸ and those isolated from the skate retina¹⁹, but hyperpolarizes them in the carp retina²⁰. In roach, GABA usually hyperpolarizes the horizontal cell membrane potential, but sometimes causes depolarization at high ($\geq 5\ \text{mM}$) concentrations. In both cases, the S-potential amplitude is reduced, and these effects are reversed by $50\ \mu\text{M}$ bicuculline, but not by $50\ \mu\text{M}$ or $250\ \mu\text{M}$ picrotoxin (Fig. 3*a*). Previous reports that picrotoxin fails to antagonize GABA have been attributed to difficulties with iontophoresis^{21,22}. Our preparation was superfused, however, and picrotoxin activity was demonstrated by its consistent antagonism of $3\ \mu\text{M}$ KA (Fig. 1*b, c*).

Fish retinal horizontal cells are postsynaptic to the dopaminergic interplexiform cells¹⁰ and exogenous dopamine (DA) modifies their electrophysiological responses²³. Therefore, we tested for interaction between DA binding sites and those which mediate the action of $3\ \mu\text{M}$ KA and QA. Dopamine at $20\ \mu\text{M}$ depolarized horizontal cells and reduced the S-potential amplitude (Fig. 3*b*) whereas $20\ \mu\text{M}$ *cis*-flupentixol, which partially blocks the DA effects, hyperpolarized the horizontal cells, and both drugs acted in the presence or absence of $3\ \mu\text{M}$ KA (Fig. 3*b, c*). Furthermore, neither $50\ \mu\text{M}$ picrotoxin nor $50\ \mu\text{M}$ bicuculline blocked the DA effects, thus we conclude that the pathways which mediate horizontal cell hyperpolarization by KA and QA are not dopaminergic.

Our data demonstrate a novel action of excitatory amino acids at the outer plexiform level of the vertebrate retina. Picrotoxin differentiates between the hyperpolarizing actions of KA and QA, but both are antagonized by bicuculline (Fig. 1*b, c*), which also blocks the effects of exogenous GABA (Fig. 3*a*). Thus, the hyperpolarizing action of both KA and QA seems to be associated with a GABAergic pathway. At low concentrations,

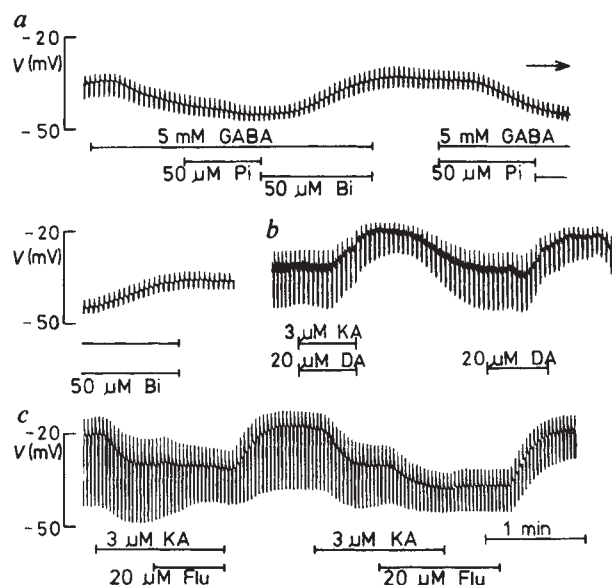


Fig. 3 *a*, Effects of GABA are reversed by $50\ \mu\text{M}$ bicuculline, but not by $50\ \mu\text{M}$ picrotoxin. *b*, Depolarization of horizontal cells by $20\ \mu\text{M}$ dopamine (DA). *c*, Hyperpolarization of horizontal cells by *cis*-flupentixol (Flu). Same protocol as for Fig. 1.

KA, QA and glutamate release GABA from isolated horizontal cells^{24,25} and there is anatomical and histochemical evidence for GABAergic feedback from horizontal cells to photoreceptors^{26,27}, which could mediate the bicuculline-sensitive hyperpolarization elicited by KA and QA (Fig. 1). Glutamate at 500 μ M hyperpolarizes fish horizontal cells¹⁸, but in roach the dose response was too variable for systematic investigation. KA- and QA-induced depolarization of retinal horizontal cells has been interpreted as evidence that photoreceptors release a glutamate-like excitatory neurotransmitter^{17,19}, but our results demonstrate that these glutamate agonists bind with greater sensitivity at sites which hyperpolarize horizontal cells through a GABA-related mechanism. Any interpretation of neurotransmission between photoreceptors and horizontal cells should take into account this hyperpolarizing pathway.

We thank H. Lundbeck & Co. for donating the *cis*-flupentixol.

Received 17 August 1983; accepted 24 January 1984.

- Shinozaki, H. & Konishi, S. *Brain Res.* **24**, 368–371 (1970).
- Johnston, G. A. R., Curtis, D. R., Davies, J. & McCulloch, R. M. *Nature* **248**, 804–805 (1974).
- McLennan, H. & Lodge, D. *Brain Res.* **169**, 83–90 (1979).
- Davies, J. & Watkins, J. S. *Brain Res.* **206**, 172–177 (1981).
- Rowe, J. S. & Ruddock, K. H. *J. Physiol., Lond.* **318**, 20P–21P (1981).
- Shiells, R. A., Falk, G. & Naghshineh, S. *Nature* **294**, 592–594 (1981).
- Curtis, D. R., Duggan, A. W., Felix, D. & Johnston, G. A. R. *Nature* **226**, 1222–1224 (1970).
- Curtis, D. R., Duggan, A. W., Felix, D. & Johnston, G. A. R. *Brain Res.* **32**, 69–96 (1971).
- Galindo, A. *Brain Res.* **14**, 763–767 (1969).
- Dowling, J. E. & Ehinger, B. *Science* **188**, 270–272 (1975).
- Cajal, S. & Ramon, Y. *Cellule* **9**, 119–253 (1893).
- Svaetichin, G. *Acta physiol. scand.* **29** (Suppl. 106), 565–600 (1953).
- Svaetichin, G. & MacNichol, E. F. Jr. *Ann. N.Y. Acad. Sci.* **74** (2), 385–404 (1958).
- Trifonov, Yu. A. *Biofizika* **13**, 809–817 (1968).
- Djamgoz, M. B. A. & Ruddock, K. H. *Neurosci. Lett.* **12**, 329–334 (1979).
- Murakami, M., Shimoda, Y., Nakatani, K., Miyachi, E. I. & Watanabe, S. I. *Jap. J. Physiol.* **32**, 927–935 (1982).
- Rowe, J. S. & Ruddock, K. H. *Neurosci. Lett.* **30**, 257–262 (1982).
- Laufer, M. *Prog. clin. biol. Res.* **113**, 257–279 (1982).
- Dowling, J. E., Lasater, E. M., van Buskirk, R. & Watling, K. J. *Vision Res.* **23**, 421–432 (1983).
- Murakami, M., Ohtsu, K. & Ohtsuka, T. *J. Physiol., Lond.* **227**, 899–913 (1972).
- Nistri, A. & Constantini, A. *Prog. Neurobiol.* **13**, 117–235 (1979).
- Davidoff, R. A. in *Epilepsy* (eds Ward, A. A. Jr, Penry, J. K. & Purpura, D.) 53–85 (Raven, New York, 1983).
- Negishi, K. & Drujan, B. D. *J. Neurosci. Res.* **4**, 311–334 (1979).
- Schwartz, E. A. *J. Physiol., Lond.* **323**, 211–227 (1982).
- Lam, D. M-K. & Ayoub, G. S. *Vision Res.* **23**, 433–444 (1983).
- Stell, W. K. & Lightfoot, D. O. *J. comp. Neurol.* **159**, 473–502 (1975).
- Lam, D. M-K., Lasater, E. M. & Naka, K-I. *Proc. natn. Acad. Sci. U.S.A.* **75**, 6310–6313 (1978).

Ganglion cell death during development of ipsilateral retino-collicular projection in golden hamster

R. Insausti*, Colin Blakemore† & W. M. Cowan

The Salk Institute for Biological Studies and
The Clayton Foundation for Research – California Division,
PO Box 85800, San Diego, California 92138, USA

* Department of Anatomy, University of Navarra, Spain

† University Laboratory of Physiology, Parks Road,
Oxford OX1 3PT, UK

In rats and hamsters all parts of the superior colliculus (SC) receive a topographically organized projection from the retina of the contralateral eye, and the rostral part also has a direct input from the lower temporal crescent of the ipsilateral retina, which views the central, binocular portion of the visual field^{1–3}. Initially the uncrossed projection covers the entire SC, but over the first 2 weeks of postnatal life it becomes progressively restricted to its adult distribution⁴. However, if the opposite eye is removed at birth there is a persistent widespread uncrossed projection to the SC^{5–8}. We have used short- and

long-term retrogradely transported neuronal markers to examine the distribution and fate of the ganglion cells of origin of the uncrossed retino-collicular projection throughout postnatal development. We conclude that the withdrawal of the early exuberant projection to the caudal SC is associated with death of ganglion cells and their virtual elimination outside the temporal crescent of the ipsilateral retina. Early enucleation of the other eye rescues many of these cells.

We performed experiments on golden hamsters (*Mesocricetus auratus*), some of which had had one eye removed by section of the optic nerve under hypothermic anaesthesia on day 0 (the first 24 h after birth). At different ages pressure injections of retrogradely transported markers (horseradish peroxidase conjugated to wheat germ agglutinin, WGA-HRP; fast blue, FB; true blue, TB; nuclear yellow, NY; and diaminodimethyl yellow, DY^{9–11}) were made through fine glass capillaries (tip diameter about 50 μ m) into the superior colliculus on one or both sides. The volume injected ranged from 15–60 nl depending on the age of the animal. For animals younger than about 6 days we used hypothermic anaesthesia and for older animals intraperitoneal chloral hydrate (3.5 mg per 10 g body weight). After periods from 13 h to 22 days the animals were deeply anaesthetized with chloral hydrate and perfused either with 0.9% saline followed by a mixture of 1% paraformaldehyde/1.25% glutaraldehyde in phosphate buffer (for the cases injected with WGA-HRP) or 1.5% saline followed by 4% paraformaldehyde in phosphate buffer (after injections of dyes). Directly after the saline perfusion, the eyes were removed and each retina was dissected out, mounted flat on a glass slide, and fixed with glutaraldehyde (for WGA-HRP) or paraformaldehyde (for the dyes). Retinae labelled with WGA-HRP were reacted on the slide by the TMB method of Mesulam^{12,13} and viewed by bright- and dark-field microscopy. Retinae containing dye tracers were washed, dried and viewed in a Leitz Dialux 20 fluorescence microscope with appropriate filters. After post-fixation and equilibration in 30% sucrose in phosphate buffer, the brains were frozen and sectioned at 50 μ m. The sections were examined either directly (for dyes) or after reacting with TMB (for WGA-HRP), to reconstruct the injection sites.

Experiments were judged successful only if (1) labelled ganglion cells were seen in the retina, with no evidence of extracellular spread of the stain to neighbouring neurones⁹, (2) the injection had adequately labelled the superior colliculus with no obvious involvement of the pretectal complex, the visual thalamus or the colliculus of the opposite side (unless that too had been deliberately injected), and (3) in the case of the enucleated animals, the entire eye and optic nerve on that side were missing. All experiments reported here were successful by these criteria.

In 87 normal hamsters, and a second group of 61 animals in which one eye had been removed on the day of birth, we injected the SC on one side with either WGA-HRP or fluorescent dye, and perfused the animals after survival times of 13–40 h, on day 1, 3, 5, 9, 13, 17, 22 or as adults. (The shorter survival periods were used for NY to minimize leakage out of labelled cells^{9,14,15}.) In some of these individual cases, especially in older animals, the injection site occupied only a small fraction of the entire SC: analysis of the resulting patterns of retinal labelling provides information about the maturation of the topography of the retino-collicular projection (which we shall report on elsewhere). Here we are concerned with the source of the early exuberant uncrossed projection to the caudal part of the SC and how its developmental elimination is affected by removal of the other eye. We, therefore, examined the labelling in the ipsilateral retinae in the 50 normal and 39 enucleated animals in which the injection site definitely involved the whole of the caudal pole and altogether covered at least the posterior two thirds of the SC.

The representative examples in Fig. 1 illustrate the way in which the distribution of ganglion cells in the ipsilateral eye (that is, the retina from which the uncrossed projection to the SC arises) progressively changes during postnatal development

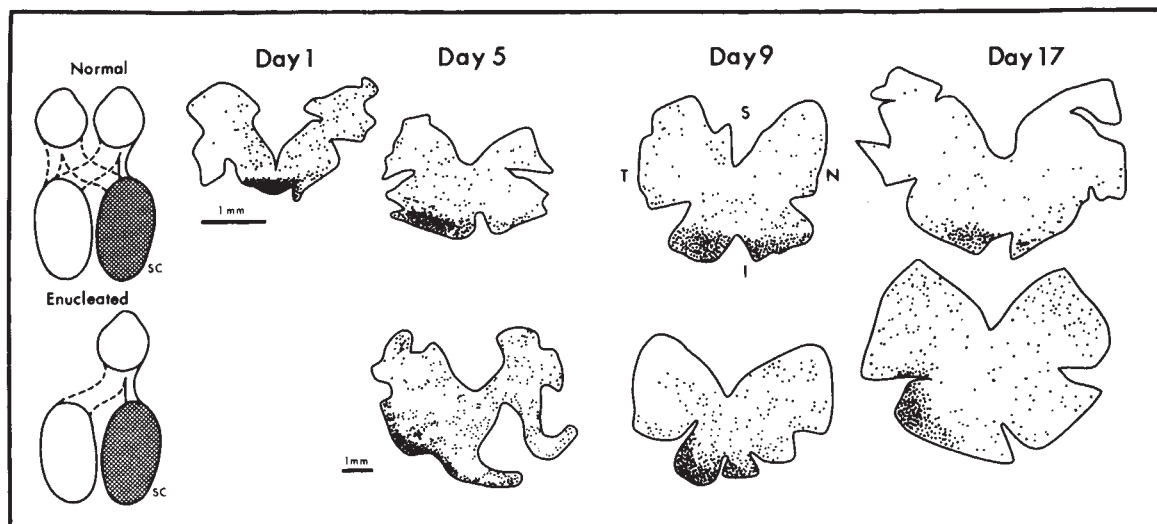


Fig. 1 Experiments to determine the distribution of ganglion cells with uncrossed retino-collicular projections at different ages. These hamsters had several injections of either WGA-HRP or fluorescent dye (see text). The cross-hatching (left) indicates the collicular injection; only in the younger animals was the rostral pole well filled. Each retina was made to lie flat by means of a number of radial cuts, the largest, along the vertical meridian of the superior retina, being used to orientate it. (N, nasal; T, temporal; S, superior; I, inferior. Note the separate calibration scale for the day 1 retina.) In every case the retina illustrated is from the right eye, ipsilateral to the injected SC. *Top*, retinæ from normal animals; *bottom*, retinæ from animals with left eyes removed on day of birth. Labelled ganglion cells indicated by dots. For the first postnatal week or so there is little difference between the two groups of animals: ganglion cells projecting to the ipsilateral SC, though especially dense in the lower temporal crescent, are scattered over the entire retina. After 9 days enucleated animals had more uncrossed cells outside the lower temporal crescent. The number of ganglion cells labelled in the temporal crescent depended on the extent to which the injection had invaded the rostral SC.

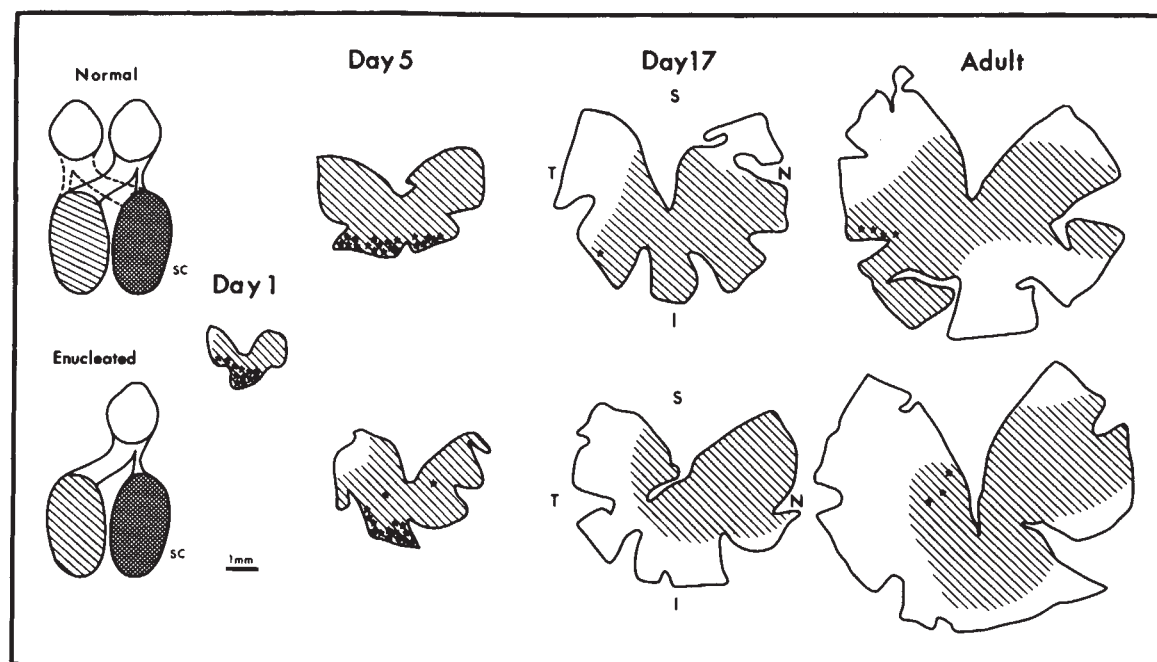


Fig. 2 Experiments to determine the number and distribution of ganglion cells with axons that bifurcate at the optic chiasma. Much of the SC on one side (always including the caudal two thirds) was injected with a blue fluorescent dye and the other side with a yellow dye. The retinæ are displayed as in Fig. 1. (All retinæ from enucleated animals were right eyes; for day 17 normal animal the left retina was analysed because it was more extensively labelled; for comparison, the retina is shown in mirror image as if from right eye so that temporal retina is always to the left.) Hatched area, dense concentration of cells labelled with dye injected into the contralateral SC: this always includes much of the nasal retina, which contributes to the early exuberant uncrossed projection. Cells labelled from the ipsilateral SC were distributed in their characteristic patterns at each age (see Fig. 1). Stars, ganglion cells unequivocally labelled with both dyes; these are more frequent in younger animals (see text), probably because the SC is smaller and more easily exposed for extensive injection during the first week of life. In some older animals the absence of intense labelling transported from the contralateral SC (hatching) in the lower temporal crescent (due to the difficulty of filling the rostral SC without involving the pretectal complex) obviously reduced our chances of detecting double-labelled cells in the temporal crescent itself. However, enucleation neither rescues nor creates a large population of cells in the nasal retina that project to both colliculi.

in normal and enucleated animals. On day 1 (injected shortly after birth) labelled cells were scattered over the entire ipsilateral retina, but with a greater concentration in the lower temporal crescent, especially when the injection site included the more rostral part of the SC. However, their overall number was low (even when the injection appeared to have filled the entire SC there were only about 200, half of which were outside the lower temporal crescent) and they were generally rather faintly labelled, perhaps because at this stage uncrossed optic axons are only just beginning to innervate the hamster SC⁴. Because of the difficulty, especially in older animals, of filling the entire SC without involving the pretectum or the opposite side of the midbrain, rigorous quantitative comparison of the total number of labelled ganglion cells at each age is not possible. However, examination of these cases in which the injection involved at least the caudal two thirds of the SC showed that, in normal animals, the number of ganglion cells labelled in the ipsilateral retina and the intensity of labelling increased over the first few postnatal days, but by day 9–13 their distribution was becoming more restricted so that fewer were seen outside the temporal crescent. By day 17 the adult pattern seemed fully established and very few cells were labelled outside the temporal crescent and the number labelled within the crescent depended on the extent to which the injection had invaded the rostral third of the SC.

By comparison, in hamsters in which one eye had been removed at birth the retinal ganglion cells with uncrossed axons remained more diffusely distributed, as in very young normal animals. After day 13, normal and enucleated animals with comparable injections differed consistently in the number of labelled ganglion cells outside the temporal crescent. Such a difference has already been described for fully-grown enucleated animals^{15,16} and our results imply that it is due largely to the partial preservation of a juvenile exuberant projection rather than the generation of an entirely novel one.

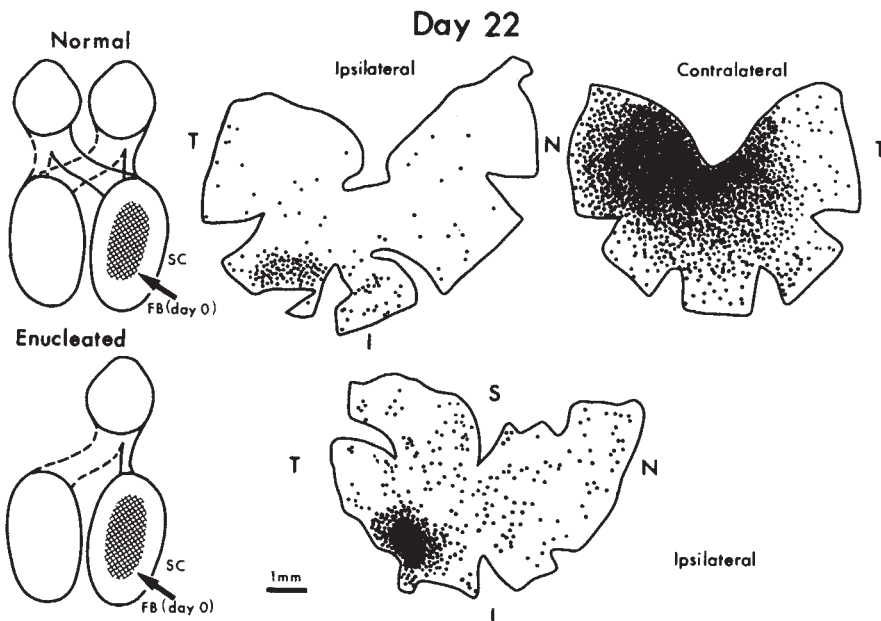
This leaves unanswered the question why enucleation of one eye at birth leads to the retention of a considerable extra

population of ipsilaterally projecting ganglion cells in the retina of the remaining eye. Cunningham's¹⁷ observations in the rat suggested that it could be due either to preservation or to genesis of numerous ganglion cells whose axons bifurcate at the optic chiasma, but this hypothesis has not been supported by double-labelling experiments in adult enucleated animals^{15,16}. To examine this idea in more detail we injected the SC with a blue fluorescent dye (FB) on one side and with a yellow dye (DY or NY) on the other in a series of 10 normal and 11 enucleated animals ranging from the day of birth to adulthood. In older animals the FB was injected some 2–4 days before the NY and the animal was perfused 18–20 h later to allow enough time for the transport of both dyes^{14,15}. Both retinas were examined carefully for cells labelled with each dye and for double-labelled neurones.

The sample results in Fig. 2 were selected because at each age the injection sites in the normal and the enucleated animal were similar in pattern, always involving the caudal part of the SC on both sides. The area of retina containing a high density of ganglion cells labelled with the dye injected into the contralateral SC is indicated by hatching: in most cases this included almost the entire nasal half. There were also many cells labelled with the dye injected into the ipsilateral SC distributed in a pattern characteristic of the age and status of the animal (compare, Fig. 1). The stars indicate the positions of cells that were labelled with *both* dyes (that is, cells with a bifurcating axon projecting to both colliculi).

Only in some of the youngest animals was the extent of the injection of each dye adequate to label almost the entire distribution of projecting ganglion cells in both eyes. Therefore, these experiments cannot provide definitive quantitative information about the total number of ganglion cells with axonal branches to both colliculi. However, comparison of normal and enucleated animals with similar injections, always involving the caudal SC, should reveal whether there is at any stage a major population of nasal ganglion cells with bifurcating axons and whether enucleation helps to preserve or generate such cells. It was clear

Fig. 3 Experiment to test whether ganglion cell death is a major factor in the withdrawal of the early, exuberant uncrossed projection, and whether such death is prevented by enucleation. Animals had the right SC extensively injected with FB on the day of birth but they were allowed to survive until day 22. In this illustration the extent of the injection site (reconstructed from sections of the brain) is drawn accurately in relation to the borders of the SC on the left: the cross-hatched areas indicate the dense staining with FB visible on day 22, though it is probable, of course, that the area of effective infiltration with FB was larger immediately after the injection on day 0. Long survival times produced a small halo of blue staining in the extracellular space and faint staining in very small, presumably glial, cells immediately around many of the labelled ganglion cells, but no evidence that label had spread significantly to other ganglion cells: the labelled ganglion cells did not occur in tightly packed clusters, which would be expected if the dye had spread from one cell to its neighbours, and the density of labelled neurones was not higher than in comparable animals injected on day 21 and killed the following day. The clear difference between the distribution in the ipsilateral nasal retina of the normal and the enucleated animal, when the injections and survival times were virtually identical, cannot be accounted for by transneuronal spread of the dye. Most of the left (contralateral) retina in the normal animal was packed with labelled ganglion cells; but the right (ipsilateral) retina had a distribution of labelled cells that was typical of a normal animal at 22 days rather than of a newborn (compare Fig. 1), implying that many cells that initially projected to the ipsilateral SC had died. On the other hand the enucleated animal had a greater density of labelled cells, even in the temporal crescent, and still had a large complement of such cells outside the crescent that had survived since the day of birth.



that even at the youngest ages the double-labelled cells were almost entirely restricted to the lower temporal crescent, just as in the normal adult rat¹⁸, and their numbers were never very high. For normal immature animals (day 13 and younger) there was an average of 13.6 double-labelled cells (S.D. = 9.36; $N = 10$ retinæ). For normal animals 17 days and older, the mean number was only 0.8 (S.D. = 1.17; $N = 10$) but the apparent decrease may be accounted for by the less complete filling of the rostral SC and hence the temporal crescent in these older animals. More important, enucleation at birth certainly did not cause a massive increase at any age in the number of double-labelled cells. Indeed, the mean number was only 5.7 (S.D. = 7.13; $N = 6$) for enucleated animals day 13 and younger and 1.2 (S.D. = 1.5; $N = 5$) for older animals. The differences in numbers between normals and enucleates were not statistically significant. It seems unlikely that cells with bifurcating axons contribute substantially to the preservation of the aberrant uncrossed projection from the nasal retina.

Finally, we tested the hypothesis¹⁵ that enucleation prevents the death of some of the ganglion cells that contribute to the exuberant uncrossed projection. In 26 normal and 16 enucleated hamsters we made injections of a long-lasting dye (FB or TB) over most of the SC on one side. The animals were then allowed to survive until days 5, 9, 13 or 22. These injections should have labelled most of the cells of origin of the early uncrossed projection and any such ganglion cells that did not die should have remained labelled when examined after the normal period of restriction of the exuberant retino-collicular projection (see refs 19, 20 and 21 for similar studies of the development of interhemispheric connections). Figure 3 shows the distribution of such neonatally-labelled ganglion cells in both retinæ of a normal hamster and in the ipsilateral retina of an enucleated animal at day 22, the longest period of survival. Comparison with the pattern of uncrossed cells in very young animals (Fig. 1) shows that the vast majority of such aberrantly projecting cells outside the temporal crescent of the ipsilateral retina have died by day 22 in the normal hamster, but many still survive in the enucleated animal. Even in the temporal crescent itself the number of labelled cells is much higher in the enucleated than in the normal animal, suggesting that throughout the retina many cells with uncrossed axons normally die but a proportion of them can be rescued by enucleation. The examples in Fig. 3 were chosen because of the almost identical appearance of their injection sites, but they are representative of the entire series in that the number of labelled cells in the ipsilateral eye was in every case higher in enucleated animals than in normals with comparable injections and survival times.

Our experiments strongly suggest that during the postnatal elimination of the early uncrossed projection to the caudal part of the SC there is widespread death of ganglion cells in the ipsilateral retina, and that neonatal removal of the other eye rescues many of these ipsilaterally projecting cells. Indeed Sengelaub and Finlay²² have observed dying ganglion cells in the retinæ of hamsters during the first 10 days of life and found that enucleation of the other eye reduces the number of dying cells²³. Recent results on cat (refs 24, 25 and D. S. Jacobs, V. H. Perry and M. J. Hawken, in preparation) and monkey²⁶ suggest that ganglion cell death during development occurs in a wide variety of mammals. Further work is needed to determine to what extent the death of ganglion cells is genetically programmed, or results from competition between optic nerve terminals from the two eyes, or between retino-tectal and cortico-tectal axons²⁷ and even between the dendrites of neighbouring ganglion cells²⁸.

This work was supported by grants from the US NIH (EY-03653), the Fundación Luis Manuel, the Sloan Foundation and the Medical Research Council. W.M.C. is a Senior Investigator of the Clayton Foundation.

Received 23 August 1983; accepted 2 February 1984.

1. Cowey, A. & Perry, V. H. *Expl Brain Res.* **35**, 457-464 (1979).
2. Frost, D. O. & Schneider, G. E. *J. comp. Neurol.* **183**, 517-568 (1979).
3. Tiao, Y.-C. & Blakemore, C. J. *comp. Neurol.* **168**, 483-504 (1976).

4. Frost, D. O., So, K.-F. & Schneider, G. E. *Neuroscience* **4**, 1649-1677 (1979).
5. Lund, R. D., Cunningham T. J. & Lund, J. S. *Brain Behav. Evol.* **8**, 51-72 (1973).
6. Land, P. W. & Lund, R. D. *Science* **205**, 698-700 (1979).
7. Thompson, I. D. *Nature* **279**, 63-66 (1979).
8. Finlay, B. L., Wilson, K. G. & Schneider, G. E. *J. comp. Neurol.* **183**, 721-740 (1979).
9. Bentivoglio, M., Kuypers, H. G. J. M. & Catsman-Berrevoets, C. E. *Neurosci. Lett.* **18**, 19-24 (1980).
10. Bentivoglio, M., Kuypers, H. G. J. M., Catsman-Berrevoets, C. E., Loewe, H. & Dann, O. *Neurosci. Lett.* **18**, 25-30 (1980).
11. Keizer, K., Kuypers, H. G. J. M., Huisman, A. M. & Dann, O. *Expl Brain Res.* **51**, 179-191 (1983).
12. Mesulam, M.-M. in *Tracing Neural Connections with Horseradish Peroxidase* (ed. Mesulam M.-M.) 1-151 (Wiley, Chichester, 1982).
13. Gerfen, C. R., Staines, W. A., Arbutnot, G. W. & Fibiger, H. C. *J. comp. Neurol.* **207**, 283-303 (1982).
14. Kuypers, H. G. J. M., Bentivoglio, M., Catsman-Berrevoets, C. E. & Bharos, T. B. *Expl Brain Res.* **40**, 383-392 (1980).
15. Jeffrey, G. & Perry, V. H. *Dev Brain Res.* **2**, 176-180 (1982).
16. Hsiao, K. & Schneider, G. E. *Soc. Neurosci. Abstr.* **6**, 684 (1980).
17. Cunningham, T. J. *Science* **194**, 857-858 (1976).
18. Jeffrey, G., Cowey, A. & Kuypers, H. G. J. M. *Expl Brain Res.* **44**, 34-40 (1981).
19. O'Leary, D. D. M., Stanfield, B. B. & Cowan, W. M. *Dev Brain Res.* **1**, 607-617 (1981).
20. Innocenti, G. M. *Science* **212**, 824-827 (1981).
21. Ivy, G. O. & Killackey, H. P. *J. Neurosci.* **2**, 735-743 (1982).
22. Sengelaub, D. R. & Finlay, B. L. *J. comp. Neurol.* **204**, 311-317 (1982).
23. Sengelaub, D. R. & Finlay, B. L. *Science* **213**, 573-574 (1981).
24. Williams, R. W., Bastiani, M. J. & Chalupa, L. M. *J. Neurosci.* **3**, 133-144 (1983).
25. Ng, A. Y. K. & Stone, J. *Dev Brain Res.* **5**, 263-271 (1982).
26. Rakic, P. & Riley, K. P. *Nature* **305**, 135-137.
27. Rhoades, R. W., Kuo, D. C. & Polcer, J. D. *Neuroscience* **7**, 2441-2458 (1982).
28. Linden, R. & Perry, V. H. *Neuroscience* **11**, 2813-2827 (1982).

Acquisition of specific cytotoxic activity by human T4⁺ T lymphocytes in culture

Bernhard Fleischer*

Institute of Virology and Immunobiology, University of Würzburg, Versbacher Str. 7, D-8700 Würzburg, FRG

Mature human T lymphocytes can be separated by monoclonal antibodies OKT4 and OKT8 according to their surface phenotypes into T4⁺T8⁻ and T4⁺T8⁺ subsets^{1,2}. From short-term experiments using bulk cultures, the helper/inducer function has been assigned to the T4⁺T8⁻ subset and the cytotoxic/suppressor function to the T4⁺T8⁺ subset^{1,2}. Thus if T lymphocytes are separated after stimulation in a mixed lymphocyte reaction (MLR), the entire cytotoxic activity is found in the T4⁺T8⁺ fraction whereas the T4⁺T8⁻ fraction shows no detectable cytotoxicity¹. If, however, T lymphocytes are cloned after MLR and grown in long-term culture, a surprisingly large fraction of T4⁺ T lymphocyte clones (TLC) shows cytotoxic activity^{3,4}. Here we report that T4⁺ TLC can acquire specific cytotoxicity during *in vitro* cultivation.

These experiments were prompted by the finding that several T4⁺T8⁻ TLC specific for influenza virus⁵ or for tuberculin⁶, which had not been cytotoxic in the initial screening after cloning, had become strongly positive for phytohaemagglutinin (PHA)-dependent cytotoxicity after several months in culture but without changing their phenotype. Several experiments were performed to prove that this cytotoxicity is indeed a property of the original TLC and is not caused by contaminating T lymphocytes derived from irradiated stimulator cells. First, as demonstrated with influenza-virus specific TLC, both proliferative response and cytotoxicity show an identical pattern of antigen specificity and HLA restriction (Table 1). Second, when the TLC are subcloned at 0.3 cells per well, all of 30 subclones show both antigen-specific proliferation and cytotoxicity (not shown). In addition, both proliferative and cytotoxic activity show the same pattern of inhibition by monoclonal anti-Ia antibodies (not shown).

To determine whether the cytotoxic activity is expressed only during phases of activation and proliferation, TLC were used in cytotoxicity assays at different times after restimulation. As

* Present address: Department of Immunology, University of Ulm, PO Box 4066, D-7900 Ulm, FRG.

Table 1 Specificity of acquired cytotoxicity is identical to the specificity of proliferative response

Target cells	% Specific lysis by clone			
	GR1	GR24	GR28	GR41
A/USSR-infected, Dw1	34	3	24	6
B/Lee-infected, Dw1	1	2	1	3
A/USSR-infected, Dw6	3	46	1	37
B/Lee-infected, Dw6	2	1	0	1
Antigen presenting cells	c.p.m. incorporated by clone			
	GR1	GR24	GR28	GR41
A/USSR-infected, Dw1	7599	370	11077	254
B/Lee-infected, Dw1	86	160	103	240
A/USSR-infected, Dw6	118	7250	280	6216
B/Lee-infected, Dw6	101	554	197	202

Four influenza A/USSR virus-specific TLC⁵ derived from donor G.R. (HLA-Dw1, Dw6) were tested simultaneously for specific cytotoxicity against infected B-lymphoblastoid (B-LCL) cells obtained from donors who shared only one HLA-D antigen with donor G.R., and for proliferative response to infected monocytes from the same donors as described⁵. Viruses used were A/USSR/90/77 and the immunologically unrelated B/Lee/40 as a control, effector: target cell ratio was 8:1 in the cytotoxicity assay.

shown in Table 2, cytotoxicity was independent of whether the TLC were freshly restimulated and still proliferating or whether the cells had returned to a resting state when the expression of Tac-1 antigen (T-cell growth factor (TCGF)-receptor)⁷ had decreased and the TLC had lost their blastoid appearance.

To follow the development of cytotoxicity 108 TLC were established from a MLR (F.J. anti-L.L.) and screened for cytotoxicity and specific proliferative response to L.L. stimulator cells as soon as sufficient cells were available. Fifteen T4⁺T8⁻ TLC that were negative for cytotoxicity and proliferated

Table 2 Expression of cytotoxicity is independent of the state of activation of the TLC

Days after last restimulation	Expression of Tac-1 antigen	% Specific lysis of target cells	
		Infected Dw1-bearing	Infected allogeneic
5	+++	32	2
22	++	42	3
31	++	47	0
38	+	37	0
45	-	38	0

Cells from TLC GR1 (specific for influenza A + HLA-Dw1) that had been restimulated with irradiated, infected HLA-matched PBL and TCGF at the intervals given were tested simultaneously for specific cytotoxicity towards A/USSR infected HLA-Dw1-bearing or allogeneic B-LCL cells. Tac-1 antigen expression was monitored with monoclonal anti-Tac antibody⁷ in indirect immunofluorescence. Effector: target cell ratio was 8:1.

strongly in response to L.L. cells were grown by restimulation with irradiated L.L. cells and tested at several intervals for antigen-specific and PHA-dependent cytotoxicity (Table 3). With all of the TLC, cytotoxicity against the stimulator cells became detectable 4 weeks after cloning and increased with extended culture. The proliferative response was unchanged. All of nine subclones, subcloned when the parental TLC were not cytotoxic, exhibited cytotoxicity when first tested approximately 3 weeks after subcloning (not shown).

To exclude the possibility that these TLC caused ⁵¹Cr-release in the absence of target cell death⁸, target cells were incubated with mitomycin C-treated TLC and labelled after 24 h with ³H-thymidine. The extent of growth inhibition was comparable to the ⁵¹Cr-release performed in parallel demonstrating that the target cells indeed had been killed (Table 4).

Table 3 Development of cytotoxicity in human T4⁺T8⁻ TLC

Clone	Days after cloning	% Specific lysis			Proliferative response (c.p.m.) to		
		L.L.	L.L. + PHA	F.J.	Medium only	L.L.	F.J.
10	18	0	4		210	12,000	160
12		0	6		168	1,995	151
27		0	6		150	33,720	80
28		0	0		106	1,116	63
48		0	5		189	22,377	107
40		40	39			NT	
10	21	0	1	1	62	15,925	105
12		2	1	0	97	12,288	139
27		0	0	0	179	19,594	123
28		0	4	0	125	12,033	130
48		0	2	0	221	11,902	238
40		55	39	3		NT	
10	28	11	17	1	306	12,114	309
12		15	16	1	262	2,309	317
27		11	23	1		NT	
28		14	21	0	281	2,119	576
48		18	23	1	166	9,060	304
40		53	43	3		NT	
10	45	4	9	0	201	6,876	252
12		17	31	0	94	2,232	137
27		18	29	0	65	1,259	64
28		14	18	1		NT	
48		18	30	0	64	1,089	58

T-cell clones were established from a secondary MLR (F.J. anti-L.L.) by limiting dilution in histoplates at 0.3 cells per well as described previously⁵. Clones were tested on days 17, 18, 21, 28, 37 and 45 after cloning for specific and lectin-dependent cytotoxicity against B-LCL cells as targets at an effector:target cell ratio of 8:1. Clones initially negative for cytotoxicity were expanded by stimulation with irradiated L.L. PBL and TCGF on days 22, 29 and 38. The clones consistently had the OKT4⁺T8⁻ phenotype. TLC no. 40 is a T4⁺T8⁻ clone used as control at an effector:target cell ratio of 2:1. NT, not tested.

Table 4 Inhibition of target cell proliferation by a cytotoxic T-cell clone

Mitomycin-C-treated GR24 cells per well	% Specific lysis		c.p.m. ³ H-thymidine incorporation by target cells (% inhibition)	
	-PHA	+PHA	-PHA	+PHA
0			23,762	17,168
2 × 10 ⁴	0	9	24,926 (-5)	12,047 (30)
4 × 10 ⁴	0	16	24,209 (-2)	8,557 (51)
8 × 10 ⁴	0	22	23,821 (0)	6,298 (63)

Mitomycin C-treated TLC were used in lectin-dependent cytotoxicity against the autologous B-LCL as target cells. A 4 h ⁵¹Cr-release assay was performed with 10⁴ labelled target cells per well in the presence or absence of PHA (10 µg ml⁻¹). For the assay of growth inhibition 10⁴ LCL cells were incubated with the TLC with or without PHA for 24 h and labelled with 0.5 µCi ³H-thymidine (2 Ci mmol⁻¹) for 16 h. Three replicates were used. The mitomycin-treated TLC incorporated less than 100 c.p.m. in TCGF.

So far, we have studied about 20 T4⁺ TLC with various antigenic specificities and all of them acquired specific cytotoxicity in culture. All required antigenic restimulation for long-term growth and did not grow in TCGF alone. The cytotoxic capacity of these clones was stable once established and was still present when the proliferative response had ceased. Therefore, only the specificity but not the function of these TLC was stable in long-term culture.

How do these findings relate to the situation *in vivo*? Evidence has been presented that at least the majority of the T4⁺ cytotoxic T lymphocytes obtained *in vitro* have specificity for class II HLA antigens^{9,10}. However, the human cytotoxic T-cell response *in vivo* to several viral infections is known to be restricted only by HLA-A or -B antigens¹¹⁻¹³, which is also true of the cytotoxicity generated in primary MLR or after *in vitro* stimulation with viral antigens¹⁴⁻¹⁶ mediated by T8⁺ T lymphocytes. On the other hand, class II HLA antigen-specific cytotoxic T lymphocytes, for example, against HLA-SB antigens can only be generated after several stimulations *in vitro*¹⁰.

The present study does not answer the question of whether T4⁺ cytotoxic T lymphocytes generally acquire their cytotoxicity only in culture, or if there exist 'true' cytotoxic T lymphocytes with specificity for class II antigens that are not detected in primary responses because of insufficient expansion *in vivo*. This study, however, clearly shows that at least some T4⁺ lymphocytes can acquire a new function while maintaining their antigen specificity.

It has recently been reported that murine T-cell clones may express multiple functions *in vitro*¹⁷ or may acquire natural-killer-like activity in culture¹⁸. Such observations have been interpreted as an indication for a functional flexibility of the immune system, capable of modifying an immune response¹⁹. However, the possibility must be considered that this flexibility is not normally realized *in vivo* but becomes apparent only *in vitro*, probably due to higher concentrations of antigen and lymphokines.

This work was supported by the DFG (SFB 105). I thank Gisela Wohlleben for assistance, Dr E. Wecker for his support and criticism and Dr T. A. Waldmann for the anti-Tac monoclonal antibody.

Received 30 September; accepted 13 December 1983.

- Reinherz, E. L., Kung, P. C., Goldstein, G. & Schlossman, S. F. *Proc. natn. Acad. Sci. U.S.A.* **76**, 4061-4066 (1979).
- Reinherz, E. L., Kung, P. C., Goldstein, G. & Schlossman, S. F. *J. Immun.* **124**, 1301-1307 (1980).
- Moretta, L. *et al. J. exp. Med.* **154**, 569-574 (1981).
- Pawelec, G., Kahle, P. & Wernet, P. *Eur. J. Immun.* **12**, 607-615 (1982).
- Fleischer, B. *Tissue Antigens* **21**, 238-245 (1983).
- Fleischer, B. & Bogdahn, U. *Clin. exp. Immun.* **52**, 38-44 (1983).
- Leonard, W. L. *et al. Nature* **300**, 267-269 (1982).
- Fischer-Lindahl, K., Nordin, A. A. & Schreier, M. *Curr. Topics Microbiol. Immun.* **100**, 1-10 (1982).
- Meuer, S. C., Schlossman, S. F. & Reinherz, E. L. *Proc. natn. Acad. Sci. U.S.A.* **79**, 4390-4395 (1982).
- Biddison, W. E., Rao, P. E., Talle, M. A., Goldstein, G. & Shaw, S. J. *exp. Med.* **156**, 1065-1076 (1982).
- Kreth, H. W., ter Meulen, V. & Eckert, G. *Med. Microbiol. Immun.* **165**, 203-214 (1979).
- Quinnan, G. V. *et al. J. Immun.* **126**, 2036-2041 (1981).
- Kreth, H. W., Kress, L., Kress, H. G., Ott, H. F. & Eckert, G. *J. Immun.* **128**, 2411-2415 (1982).

- McMichael, A. J. *Springer Semin. Immunopath.* **3**, 3-22 (1980).
- Wallace, L. E., Moss, D. J., Rickinson, A. B., McMichael, A. J. & Epstein, M. A. *Eur. J. Immun.* **11**, 694-699 (1981).
- Kress, H. G. & Kreth, H. W. *J. Immun.* **129**, 844-849 (1982).
- Dennert, G., Weiss, S. & Warner, J. F. *Proc. natn. Acad. Sci. U.S.A.* **78**, 4540-4543 (1981).
- Brooks, C. *Nature* **305**, 155-158 (1983).
- Pawelec, G., Schneider, E. M. & Wernet, P. *Immun. Today*, **4**, 275-278 (1983).

Aged murine killer T-cell clones acquire specific cytotoxicity for P815 mastocytoma cells

M. M. Simon, H. U. Weltzien, H. J. Bühring & K. Eichmann

Max-Planck-Institute for Immunobiology, Freiburg, FRG

T-cell clones that grow continuously in tissue culture have become a major tool for studying the properties of T lymphocytes. It is therefore important to know to what extent such clones resemble their normal counterparts. Several reports have appeared recently which demonstrate that long-term T-cell lines may lose the specificity for which they were initially selected and acquire cytotoxic activity to a variety of targets, typical of the activity displayed by natural killer cells¹⁻³. We now report a number of instances in which murine cytotoxic T-cell clones have lost their original specific cytotoxic activity but have acquired strong specific cytotoxic activity for P815 mastocytoma target cells. Loss of the original specificity was usually observed after continuous *in vitro* cultivation for more than 6 months. We propose that this novel type of cytotoxicity should be called aged killer activity.

We present data from three types of cloned cytotoxic T cells. First, we show results from a number of cytotoxic lymphocyte (CTL) clones derived from female C57BL/6 (B6) mice with specificity for male B6 target cells carrying the H-Y antigen. These clones were obtained by standard procedures^{4,5} (see Fig. 1 legend) and maintained in the presence of irradiated spleen cells from male B6 mice (stimulator cells) and the supernatant from cells stimulated with concanavalin A (Con A-SN). After loss of H-Y specific cytotoxic activity, maintenance in Con A-SN without stimulator cells is usually sufficient. Second, we show data obtained with CTL clones with specificity for trinitrophenyl (TNP) modified syngeneic target cells. These clones are derived from B6 spleen cells undergoing a primary response *in vitro* to trinitrophenylated syngeneic cells. After three successive restimulations in the presence of Con A-SN the cells were cloned and maintained in the same way as the B6 anti-H-Y clones. Third, we present experiments with a murine T-cell line (clone 96) obtained in 1980 and kindly given to us by Dr P. Krammer, German Cancer Research Centre, Heidelberg. This cell line, of unknown origin, has never been known to possess any antigen-specific cytotoxic activity and is used as an indicator for interleukin-2 (IL-2) activity. All the CTL clones described here are positive for Thy 1 and Lyt 2 antigens.

Fig. 1 Cloned CTLs lose their original specific cytotoxic activity against H-Y and acquire P815-specific cytotoxicity. *a*, Specificities and cytotoxic activities of an anti-H-Y T-cell line and two recently derived anti-H-Y T-cell clones. *b*, Comparison between the specificities of a two-month-old H-Y specific T-cell clone (2D10.2) and four aged (6–8 months) H-Y specific T-cell clones (2A5, 2C5, 1E3, 3E6).

Methods: For the establishment of long-term H-Y specific T-cell lines 2.5×10^7 spleen cells from female C57BL/6 mice (B6), previously primed to male B6 cells by i.p. injection of 3×10^7 B6 spleen cells, were cultivated *in vitro* in 50 ml tissue culture flasks (Nunc, 163371) and restimulated at weekly intervals with 2.5×10^7 irradiated male B6 stimulator cells in RPMI 1640 medium supplemented with 10% fetal calf serum, 10^{-5} M 2-mercaptoethanol, 2.5 mM HEPES, 2 mM glutamine, Kanamycin ($100 \mu\text{g ml}^{-1}$) and tyrosine ($10 \mu\text{g ml}^{-1}$ in a 5% CO_2 atmosphere. After 6 weeks the anti-H-Y specific T-cell line was maintained and expanded in 50 ml tissue culture flasks on antigen in supernatant from rat spleen cells sensitized with $5 \mu\text{g ml}^{-1}$ concanavalin A in medium for 24 h (Con A-SN). H-Y specific CTL clones were obtained using limiting dilution procedures. Responder cells were seeded from long-term cultures at 10–0.3 cells per well and expanded on 5×10^5 irradiated male B6 stimulator cells in Con A-SN in round-bottom microtitre plates (Nunc, 1-63320). After 2–3 weeks cytotoxic activity was tested in individual wells on ^{51}Cr -labelled Con A stimulated male B6 spleen cells and the positive lines were picked and expanded on antigen in the presence of Con A-SN in a volume of 2 ml in microtitre plates (Linbro, 1624 TC) and subsequently in 50 ml tissue culture flasks under similar conditions. CTL lines 3E6 and E1 were recloned twice on antigen and Con A-SN at dilutions of 0.3 cells per well under similar conditions. Lymphocytes were assayed in a 4 h ^{51}Cr -release cytotoxic assay at various effector to target cell ratios for their ability to lyse various target cells. Effector cells in 0.1 ml of culture medium (RPMI 1640, supplemented with 10% fetal calf serum and 0.5 mM HEPES) were added to 2×10^3 ^{51}Cr -labelled Con A stimulated spleen cells of male or female B6 mice or to ^{51}Cr -labelled tumour targets P815 (H-2^d) and YAC (H-2^a). To test for lectin-mediated lympholysis, effector cells and ^{51}Cr -labelled P815 target cells were incubated in the presence of PHA (Gibco, 670/0576, 4% of stock solution). After incubation for 4 h at 37 °C the plates were centrifuged at 1,000 r.p.m. for 10 min and 0.1 ml supernatant was removed for counting. Percent specific lysis was calculated as $100 \times (\text{experimental-low control}) / (\text{acid lysis-low control})$. Each effector cell concentration was tested in triplicate. Non tumour cell targets (B6, B10.D2, DBA/2, AKR) were Con A blasts. All targets were tested at all effector to target (E:T) ratios; negative lysis is indicated for highest E:T ratio only.

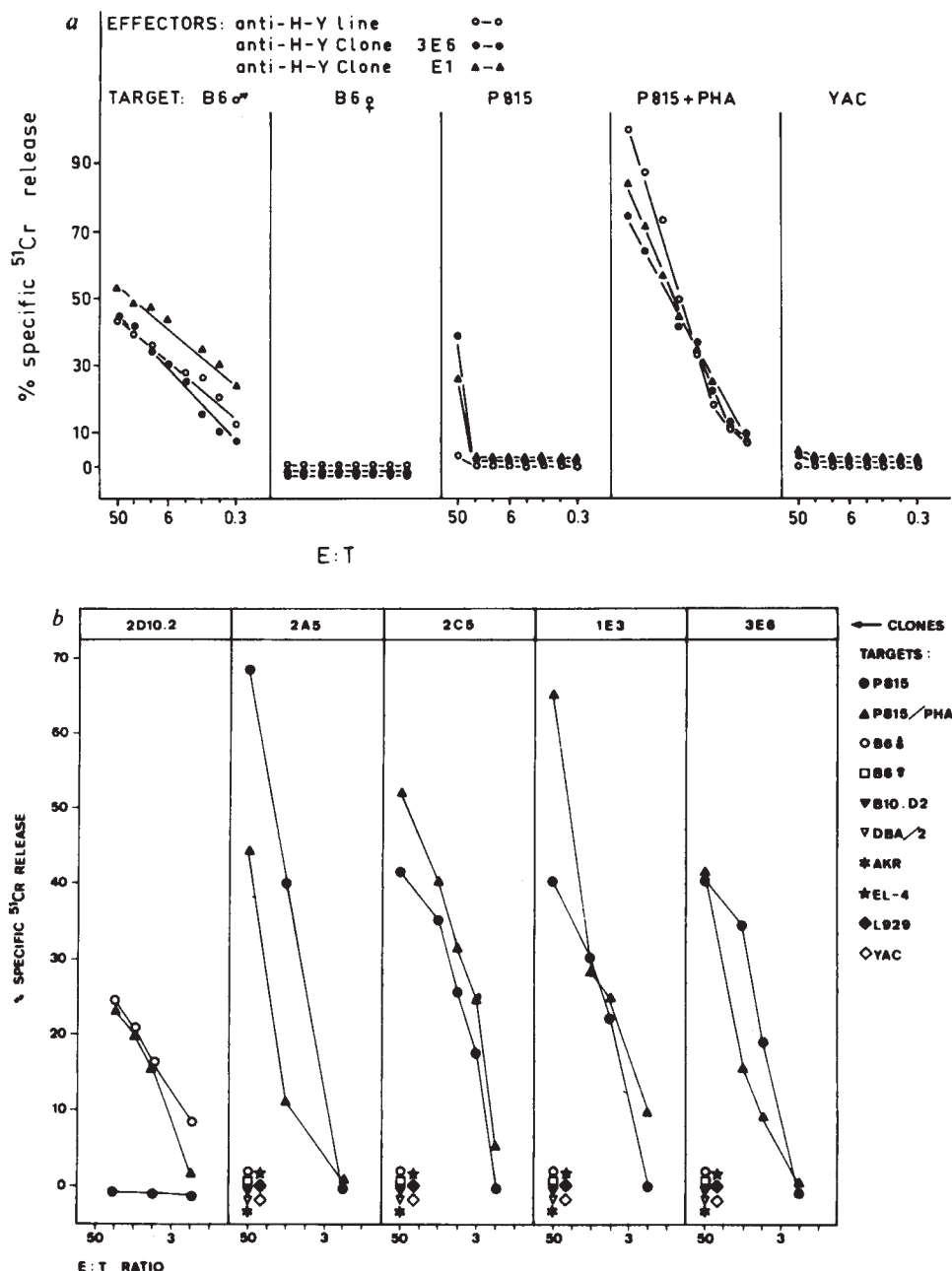


Figure 1a shows the cytotoxic activity and specificity of an uncloned anti-H-Y T-cell line and two recently derived anti-H-Y CTL clones. The data show that cytotoxicity is largely specific for male B6 target cells. Whereas only marginal cytotoxicity is observed on P815 cells, strong lectin (phytohaemagglutinin, PHA) facilitated lysis of this target is seen. Cytolytic activity on YAC tumour cells is totally absent although this target has been shown in other experiments to be readily lysed by natural killer (NK) cells or alloreactive CTL. After 5–7 months of continuous *in vitro* culture, this H-Y specific cytotoxicity of the clones is lost. Uncloned CTL usually maintain their original activity for somewhat longer periods.

The data in Fig. 1b demonstrate aged killer (AK) activity for P815 of four B6 anti-H-Y clones after more than 7 months of *in vitro* culture. For comparison, one recently derived B6 anti-H-Y clone (2D10.2) was included in the experiment. Whereas clone 2D10.2 shows H-Y specific killing and lectin facilitated kill on P815 but no AK activity, all aged clones kill P815 cells without any remaining activity for B6 male target cells. The total loss of specific cytotoxicity in AK cells was further demonstrated in cold target inhibition experiments in which P815 cells but not B6 male target cells were active as inhibitors (data not shown). The data show also that AK activity is rarely enhanced and often somewhat depressed by the addition of PHA. The

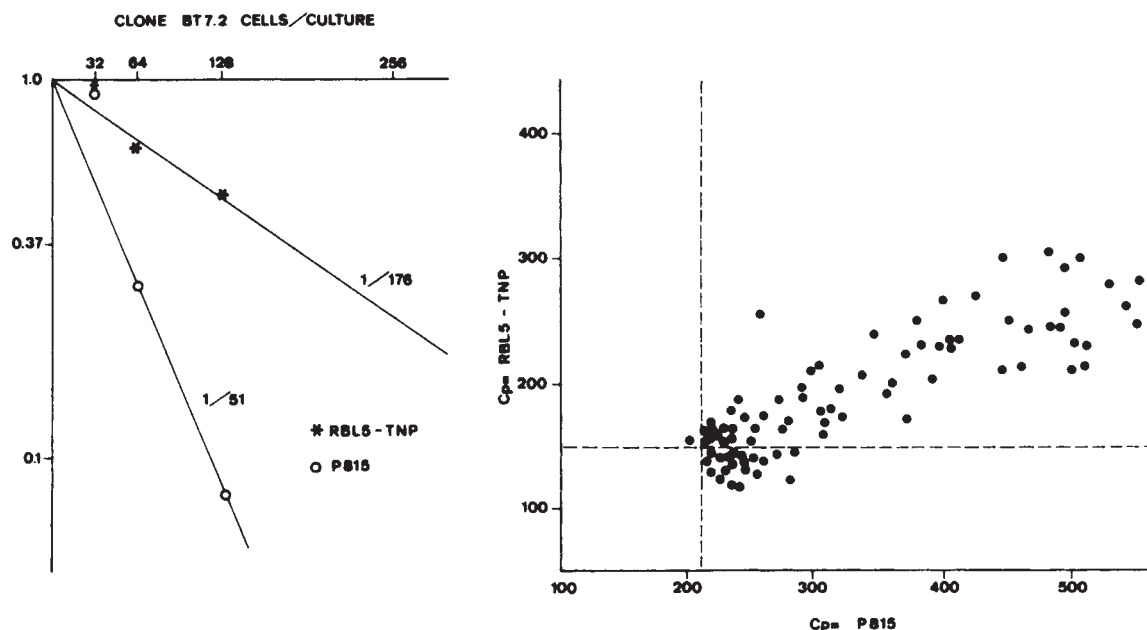


Fig. 2 Cytotoxic activity of the TNP-specific, H-2^b restricted B6-CTL clone BT7.2 4.5 months after induction and 3 months after subcloning. Ficoll-purified cloned cells were seeded in limiting dilution (24 wells per group) on 5×10^5 irradiated B6-TNP spleen cells and T-cell growth factor (TCGF) in 10% FCS containing RPMI-medium in round-bottom microtitre culture plates. On day 6, cultures were split and assayed in parallel on ^{51}Cr -labelled P815 and RBL5-TNP target cells (2×10^3 per well). Cultures were considered positive when lysis exceeded the mean plus 3 standard deviations of 24 control wells containing all ingredients except clone BT7.2 cells. The left panel shows the fraction of negative wells per group versus the number of BT7.2 cells per well. Frequency determinations were done by minimal χ^2 method. The right panel represents chromium release in individual wells on the two respective targets and includes only those wells with positive kill on at least one of the two targets. Dotted lines indicate the mean chromium release of the control groups plus 3 standard deviations.

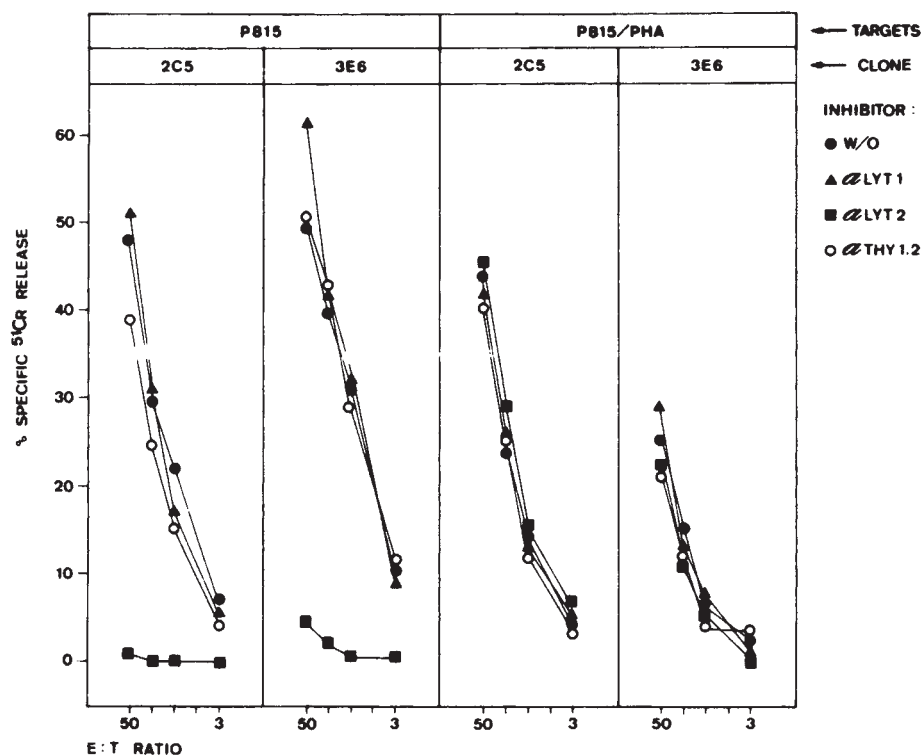


Fig. 3 Blocking of cytotoxicity of aged killer cells by monoclonal anti-Lyt 1, anti-Lyt 2 and anti-Thy 1.2 antibodies. Aliquots of 1.2×10^6 cloned T cells of lines 2C5 and 3E6 were preincubated for 30 min at room temperature with 100 μl supernatant fluids from rat hybridoma cell lines producing monoclonal antibodies directed against non-polymorphic determinants of Lyt 1 (53-7.3) or Lyt 2 (56-6.7) or with 100 μl ascites fluid (dilution 1:200) containing monoclonal anti-Thy 1.2 antibodies (Code 5B5 Olac). Cell suspensions were then diluted and assayed in a short-term ^{51}Cr -release cytotoxicity assay at various effector to target cell ratios for their ability to lyse P815 tumour target cells in the absence (left) or presence (right) of PHA. Experimental conditions as well as calculation for per cent specific lysis were as described in Fig. 1.

specificity of AK activity and its distinctness from NK activity is shown by the failure to lyse eight other target cells, including the NK sensitive targets YAC and EL4. Note that the failure of aged CTL clones to lyse H-2^d targets shows that AK activity is not caused by the frequently observed cross-reactivity of anti-H-Y clones with H-2D^d (refs 6, 7). It was also established, by post-labelling procedures (data not shown), that the release of ^{51}Cr from P815 target cells was accompanied by cell death and not due to cytolytic effects⁸.

Similar observations were made with aged B6 anti-B6-TNP clones. Specific cytotoxicity for TNP-coupled syngeneic cells

and PHA facilitated kill on P815 was observed in recently derived clones, whereas AK activity was observed several months later, usually before the clones had lost their original specific cytotoxicity. For example, clone BT7.2 had both activities after 4 months of culture. The experiment in Fig. 2 was performed to see whether both cytotoxicities are mutually exclusive or can be expressed by the same cell. Limiting dilution analysis of clone BT7.2 revealed the frequency of P815 killing cells to be 3 times greater than that of cells killing TNP coupled syngeneic cells. However, the results of split cultures show that both killing activities are highly correlated, strongly suggesting

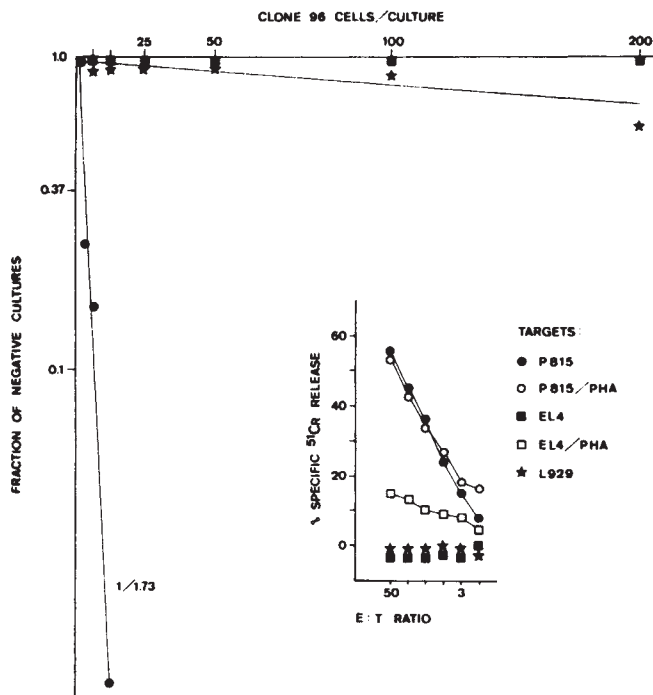


Fig. 4 Limiting dilution analysis of cells killing P815 (H-2^d), EL4 (H-2^b), and L929 (H-2^k) target cells in clone 96. Clone 96 cells were cultured (24 wells per cell) in 20% Con A-SN in the presence of 10⁵ irradiated strain AKR (H-2^k) spleen cells for 7 days. Thereafter cultures were assayed for cytotoxicity as described in Fig. 1. Cultures were considered positive when lysis exceeded the mean plus 3 standard deviations of 24 control wells containing all ingredients except clone 96 cells. Frequency determination according to minimal χ^2 method. Insert: bulk culture cytotoxicity of clone 96 for various targets.

that most cells express both activities at the same time. This suggests cells gradually change from specific cytotoxicity to AK activity.

Inhibition of cytotoxicity by anti-Lyt 2 antibody is characteristic for many but not all CTL^{9,10} although the molecular basis for this inhibition is still unknown. Typical NK cells and Con A activated large granular lymphocytes with multiple specificities are not inhibited by anti-Lyt 2 antibody^{2,11}. The data in Fig. 3 show that AK activity is readily and specifically inhibited with monoclonal anti-Lyt 2 antibody, and that inhibition can be abrogated by addition of PHA. This important similarity between specific cytotoxicity and AK activity suggests that the latter involves a specific recognition process similar to that of antigen recognition.

In studies with long-term tissue culture lines, particularly if they dramatically change their properties, the possibility of accidental contamination with other cell lines has to be considered. We have examined the possibility of contamination by limiting dilution and cloning experiments and have found that practically every cell in our aged CTL clones possesses AK activity. A limiting dilution experiment (Fig. 4) shows that the frequency of cells with cytotoxic activity for P815 in clone 96 is 1/1.7. The frequencies of specific CTL in recently derived CTL clones, determined by similar methods, are rarely greater than 1/10 (M.M.S. and K.E., unpublished data). In a cloning experiment, more than 90% of 110 subclones of clone 96 had detectable P815 cytotoxicity suggesting that clone 96 is homogeneous with respect to AK activity (data not shown). We have also considered the possibility that AK activity is the result of some infection, possibly with mycoplasma in our long-term tissue culture lines. This is unlikely because, on the one hand, mycoplasma has not been found in our cell lines and, on the other hand, many cell lines, including tumour cells and helper T-cell clones, are kept in our laboratory under conditions similar

to that used to culture CTL clones and none of these lines show AK activity.

Taken together, we think that the acquisition of AK activity by aged CTL clones is a frequent phenomenon that occurs independent of the original antigen specificity for which clones were selected. It appears to be a gradual change within each cell so that both cytotoxic activities may transiently be expressed at the same time. We consider these results of significance for several reasons. First, a fair proportion of our knowledge of murine cytotoxic T cells comes from experiments in which CTL from B6 mice with specificity for H-2^d alloantigens are assayed on P815 (H-2^d) cells. In the future, when long-term *in vitro* cultured CTL are used, H-2^d specific cytotoxicity must be distinguished from AK activity. Second, since AK activity seems to involve partially the same molecular apparatus as specific cytotoxicity its study may furnish some insight into the mechanisms of specific cytotoxicity.

Received 15 July 1983; accepted 5 January 1984.

1. Hengartner, H. *et al.* in *NK cells and Other Natural Effector cells* (ed. Herberman, R. E.) 893-902 (Academic, New York, 1982).
2. Brooks, C. G., Karibojashi, K., Sale, G. E. & Henney, C. J. *Immun.* **128**, 2326-2335 (1982).
3. Binz, H., Fenner, M., Frei, D. & Wigzell, H. *J. exp. Med.* **157**, 1252-1260 (1983).
4. v. Boehmer, H., Haas, W., Pohl, H., Hengartner, H. & Nabholz, M. *Springer's Sem. Immunopath.* **3**, 23-49 (1980).
5. Glasebrook, A. L. *et al. Immun. Rev.* **54**, 225-266 (1981).
6. v. Boehmer, H. *et al. Eur. J. Immun.* **9**, 592-597 (1979).
7. Kanagawa, O., Louis, J. & Cerottini, J. C. *J. Immun.* **128**, 2362-2366 (1982).
8. Lindahl, K. F., Nordin, A. A. & Schreier, M. H. in *T Cell Hybridomas* (eds v. Boehmer, H., Haas, W., Köhler, G., Melchers, F. & Zeuthen, M.) 1-10 (Springer, Berlin, 1982).
9. Nakayama, E., Shiku, H., Stockert, E., Oettgen, H. F. & Old, L. *Proc. natn. Acad. Sci. U.S.A.* **76**, 1977-1981 (1979).
10. Möller, G. (ed.) *Immun. Rev.* **68**, 5-218 (1982).
11. Shortman, K., Wilson, A., Scollay, R. & Chen, W. F. *Proc. natn. Acad. Sci. U.S.A.* **80**, 2728-2732 (1983).

Evolution of a trypanosome surface antigen gene repertoire linked to non-duplicative gene activation

M. Laurent[‡], E. Pays*, K. Delinte*, E. Magnus†, N. Van Meirvenne† & M. Steinert*

* Département de Biologie Moléculaire, Université libre de Bruxelles, Rhode St Genèse, Belgium

† Laboratorium voor Serologie, Instituut voor tropische Geneeskunde, Antwerp, Belgium

African trypanosomes activate, one at a time, a large set of genes coding for different variant-specific surface antigens (VSAs). These genes have been classed into two groups. In the first group a permanently silent basic gene copy is duplicated and the expression-linked copy (ELC) transposed to an expression site located at a chromosome end¹⁻³. The process is a gene conversion which changes a variable stretch of the preceding ELC⁴. Genes belonging to the second group do not give rise to an additional copy when expressed by a still unknown mechanism^{5,6}. We report here that the gene for antigenic type AnTat 1.6 is located in a telomeric DNA region and is expressed without being duplicated. In clone AnTat 1.6 and the ensuing ones, the ELC of the preceding VSA (AnTat 1.3) is conserved, but in an inactive conformation. Moreover, the AnTat 1.6 gene is lost from the genome of the AnTat 1.6-derived variants, in which the duplication-linked mechanism of gene activation occurs: the gene appears to be replaced by the incoming ELC. These observations show that a trypanosome surface antigen repertoire may evolve by loss and gain of VSA genes, depending on the alternation of the different recombinational mechanisms involved in antigenic variation.

All the VSA genes studied so far in the AnTAR 1 repertoire of *Trypanosoma brucei brucei* (stock EATRO 1125) are activated by the duplicative transposition process: these genes are

‡ To whom reprint requests should be addressed.

Clone derivation (AnTat)
stock EATRO 1125

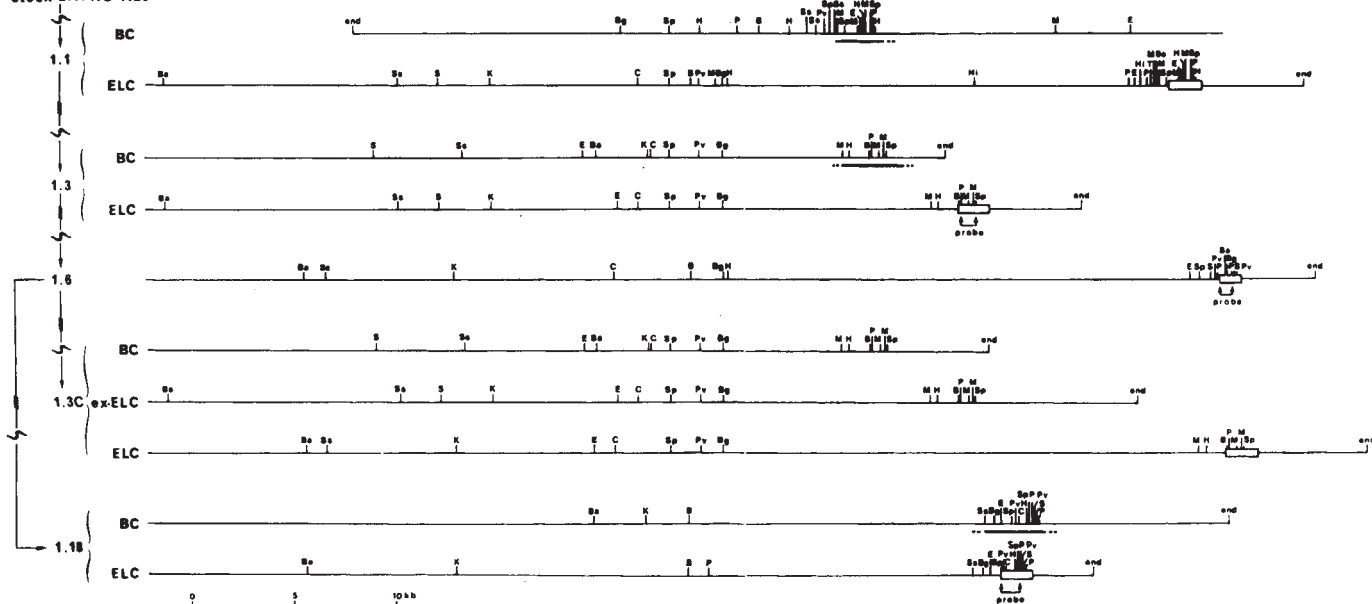


Fig. 1 Restriction maps of the telomeres containing the AnTat 1.1, 1.3, 1.3 C, 1.18 BCs and ELCs, as well as the AnTat 1.6 gene. The bars under the BC maps represent the known extent of the transposed sequences, with uncertainties at both ends. The sequences corresponding to the known cDNAs are represented by boxes, with the fragments used as probes indicated by arrows. An abridged pedigree of the clones is shown on the left: starting from the EATRO 1125 stock, VATs were successively isolated by immunological selection and cloning²⁰. ~, switch of antigenic type; ~, cloning; abbreviations: B, *Bgl*I; Ba, *Bam*HI; Bg, *Bgl*II; C, *Cla*I; E, *Eco*RI; H, *Hind*III; Hi, *Hin*fI; K, *Kpn*I; M, *Msp*I; P, *Pst*I; Pv, *Pvu*II; S, *Sal*I; Sa, *Sau*3A; Sa96I, *Sau*96I; Sp, *Sph*I; Ss, *Sst*I; T, *Taq*I.

those determining the antigenic types AnTat 1.1 and 1.8 (ref. 2), 1.3 (ref. 7), 1.13 (ref. 8) and 1.30 (our unpublished results). The gene corresponding to the VAT AnTat 1.3, a protein which always appears early in a chronic infection, is remarkable in that its basic copy (BC) is located in a telomeric DNA region which, from restriction mapping, appears almost identical to the expression site⁷. This peculiarity prompted us to speculate that such 'telomeric' VSA genes could possibly be exchanged with the corresponding sequence in the expression site by a reciprocal crossing-over; provided that the cross-over point occurs between the gene and the putative upstream promoter region, the gene would be activated without giving rise to an additional copy and a single promoter could account for the activation of all the VSA genes⁷. This model⁹ is supported by the observation that the VSA genes expressed without an ELC are telomeric¹⁰⁻¹².

The model implies that the ELC of the gene expressed before the crossing-over will stay as a new silent member of a multigene family. Most of the VSA genes examined so far belong to such multigene families^{2,11,12}. Conversely, it can be predicted that the gene which has been activated by the reciprocal telomeric recombination will be lost if the following antigenic switch is achieved by gene conversion. We therefore investigated whether such loss and gain of VSA genes accompanied antigenic variation in the following sequence of cloned populations:

AnTat 1.1 → AnTat 1.3 → AnTat 1.6 → AnTat 1.3C
AnTat 1.18

Poly(A)⁺ RNAs were isolated from these clones and used as templates for the synthesis of double-stranded complementary DNAs (ds cDNAs) which were cloned in the plasmid pBR322 by the standard procedure^{7,13}. AnTat 1.3, 1.6 and 1.18 specific sequences were selected by double screening¹³ and their specificity was checked by hybridization to Northern blots of poly(A)⁺ RNAs from various clones. The cloned sequences were then used as ³²P-labelled probes to analyse the corresponding genes by hybridization with Southern blots¹⁴ of genomic DNA digests from the various trypanosome clones.

An analysis of the AnTat 1.3 gene has been reported previously⁷; Fig. 1 presents an updated restriction map of this gene

and its environment. That the AnTat 1.3 gene is activated by duplicative transposition is confirmed here by the presence of an additional copy of the gene in the AnTat 1.3 clone as compared with the preceding AnTat 1.1 clone (Fig. 2). Surprisingly, we observed that this ELC, present in a 14.9 kilobase (kb) *Sph*I fragment (Fig. 2), is conserved in the ensuing AnTat 1.6, 1.3C and 1.18 clones. Since this copy is the one transcribed in the AnTat 1.3 genome⁷, we investigated whether it could still be active in the following AnTat 1.6 variant. Therefore, AnTat 1.3 and 1.6 nuclei were submitted to a mild DNase I hydrolysis, to probe the chromatin conformation around the AnTat 1.3 ELC sequence in both variants; actively transcribed DNA is indeed more sensitive to DNase I attack than inactive DNA¹⁵. Figure 2 (*Sph*I digest) suggests that the AnTat 1.3 ELC staying in AnTat 1.6 DNA (ex-ELC) has lost much of its DNase I sensitivity, suggesting that it is no longer transcribed. On the other hand, a new additional copy was revealed in the second AnTat 1.3 expressor clone (AnTat 1.3C): this new ELC (in a 28-kb *Sph*I fragment, Fig. 2) seems to be the one transcribed, as indicated by its high DNase I sensitivity (Fig. 2).

The restriction map of the AnTat 1.6 gene and adjacent DNA collected using AnTat 1.6 cDNA gene probes, is also presented in Fig. 1. Like the AnTat 1.3 gene⁷, the 1.6 gene is telomeric, as indicated by its susceptibility to exonuclease *Bal*31 digestion (Fig. 3) and by the apparent clustering of all restriction sites at the 3' end. However, it is not duplicated when expressed: no additional copy has ever been found, as shown in Fig. 3 after *Bgl*II-*Eco*RI digestion and in other digests (results not shown). Moreover, as shown by DNase I sensitivity tests (Fig. 3) both the 1,200-base pair (bp) *Bgl*II-*Eco*RI fragment and the *Eco*RI 4.4-kb telomeric fragments, which restriction analyses indicate as clearly derived from the AnTat 1.6 gene, are in an inactive chromatin conformation in the AnTat 1.3 clone but appear DNase I hypersensitive in AnTat 1.6 nuclei. We conclude that the same sequence which is silent in clone AnTat 1.3, is activated in the AnTat 1.6 genome without being duplicated. In contrast with the AnTat 1.3 ELC, the AnTat 1.6 gene is lost in the ensuing variants AnTat 1.3C and AnTat 1.18 (Fig. 3: *Bgl*II-*Eco*RI digests).

These observations can be easily explained if we suppose that, in the AnTat 1.6 genome, the telomere harbouring the AnTat

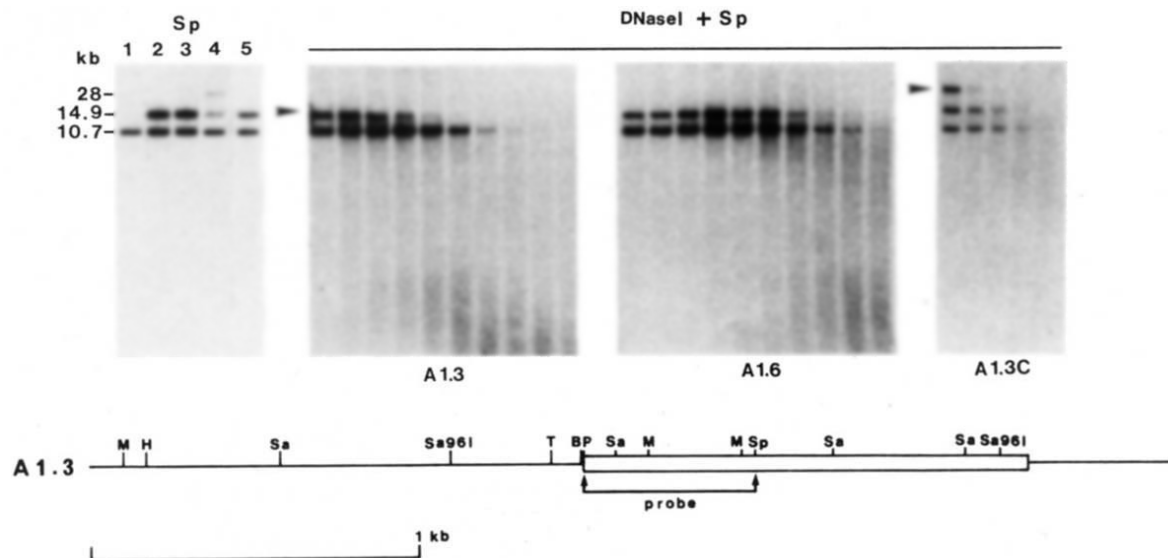


Fig. 2 Fate of the AnTat 1.3 ELC in different clones. The first panel shows the *Sph*I fragments in AnTat 1.1, 1.3, 1.6, 1.3C and 1.18 genomic DNAs, in tracks 1 to 5 respectively. The genomic DNAs were extracted from the different trypanosome clones as described previously², digested by restriction endonuclease, electrophoresed on 0.85% agarose gels, transferred to nitrocellulose and hybridized with ³²P-labelled *Pst*I-*Sph*I fragment of AnTat 1.3 cDNA (see map below). The three other panels show the AnTat 1.3 gene sensitivity to DNase I digestion in nuclei from clones AnTat 1.3, 1.6 and 1.3C respectively. The nuclei, prepared according to Pays *et al.*²¹, were digested by DNase I for 0, 7.5, 15, 30, 60, 120, 240, 360, 480 and 600 s for AnTat 1.3 and 1.6, and for 0, 10, 20, 30 and 240 s for AnTat 1.3C. The DNA was extracted and analysed as indicated above. Arrows point to the sequences preferentially digested by DNase I.

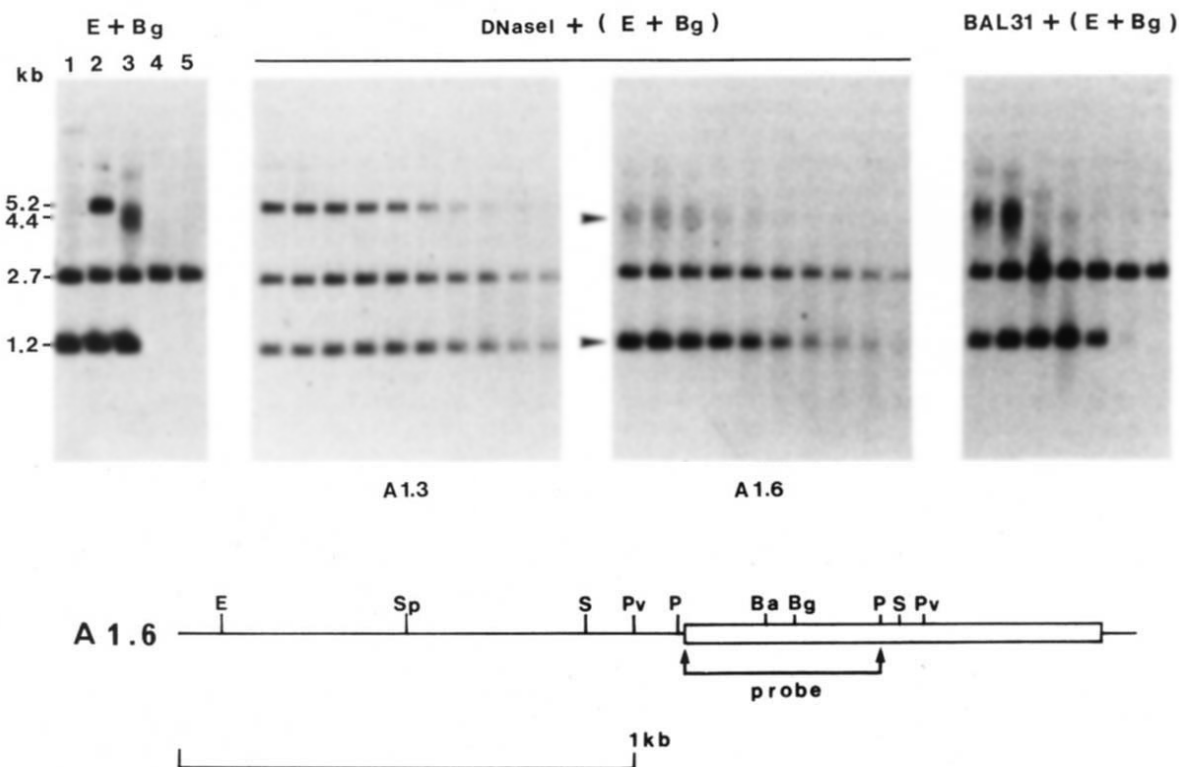


Fig. 3 Analysis of the AnTat 1.6 gene in different trypanosome clones. The first panel shows the AnTat 1.6-specific *Eco*RI-*Bgl*II fragments in AnTat 1.1, 1.3, 1.6, 1.3C and 1.18 genomic DNAs, respectively in tracks 1-5. Method as for Fig. 2, but the blots were hybridized with an AnTat 1.6 probe (see restriction map in this figure). The two next panels show the AnTat 1.6 gene sensitivity to DNase I digestion in AnTat 1.3 and 1.6 nuclei. The assays were carried out as described in Fig. 2 legend, but the final digestion was by *Eco*RI-*Bgl*II, and the probe was from AnTat 1.6. Arrows point to the sequences preferentially sensitive to DNase I, namely, the 1.2-kb internal fragment of the gene and the 3' end of the AnTat 1.6 telomere. The last panel shows that the AnTat 1.6 gene is telomeric: a *Bal*31 exonuclease digest has been performed on AnTat 1.6 genomic DNA for 0, 5, 10, 15, 20, 30 and 45 min from left to right, as previously described⁷. The DNA was then hydrolysed by *Eco*RI and *Bgl*II and hybridized with the AnTat 1.6 probe.

1.3 ELC has been exchanged by reciprocal recombination with the telomere containing the AnTat 1.6 gene. Indeed, provided the crossing-over occurs between the ELC and the transcription promoter, such rearrangement would activate the AnTat 1.6 gene while turning off the AnTat 1.3 ELC: the latter will then appear as a new member of the nascent AnTat 1.3 gene family, whereas the AnTat 1.6 gene will be chased from the next variant genome and replaced by the following ELC. This hypothesis is strongly supported by the following observation. The telomeres harbouring the AnTat 1.3C and 1.18 ELCs, two variants derived from AnTat 1.6, have restriction maps similar, perhaps identical, to the telomere containing the AnTat 1.6 gene, taking into account that, with the available probes, all the restriction sites could not be mapped in each case (Fig. 1). This suggests that the AnTat 1.6 gene has been converted by the AnTat 1.3C or 1.18 BCs within its own telomere. Like the expression site described so far for several variants^{7,8}, the telomere harbouring the AnTat 1.6 gene can be a target for gene conversion. Apparently, only 'activated' telomeric sequences become a target for gene conversion, at least at the frequency generally observed for antigenic-type switching. Reciprocal recombination has also been proposed as the most probable mechanism for the activation of the AnTat 1.1C gene¹⁶.

Two other models could account for the non-duplicative activation of the AnTat 1.6 gene: (1) a transcription promoter upstream from the gene has been activated in the AnTat 1.6 clone, while the previously active AnTat 1.3 promoter has been simultaneously silenced; (2) a mobile promoter has been transposed from the telomere containing the AnTat 1.3 ELC to that containing the AnTat 1.6 gene. Although neither of these two hypotheses can presently be ruled out, we favour the straightforwardness of the telomere exchange model.

Whatever the AnTat 1.6 gene activation mechanism may be, it clearly leads to the appearance of a new AnTat 1.3 related sequence in the genome of AnTat 1.6 and of the ensuing variants. This sequence is the previously expressed AnTat 1.3 ELC, which is conserved in the next clones in an inactive chromatin conformation. From the AnTat 1.6 clone onwards, the AnTat 1.3 sequences thus belong to a multigene family. This family is increased by one additional member in the AnTat 1.3 clone derived from AnTat 1.6 (namely AnTat 1.3C). Alternation of duplicative transposition and nonduplicative activation can thus clearly lead to the generation and development of antigen gene families. As different members of the same family may diverge in such a way that different antigenic types will be expressed, as shown for instance in the case of AnTat 1.10⁴, this alternation must actively participate in the evolution of antigen repertoires.

In contrast with AnTat 1.3 gene addition, the AnTat 1.6 gene is lost from the ensuing variants, for instance AnTat 1.3C, 1.18 and also 1.16 (results not shown for the latter). This has been confirmed by serological examination: no AnTat 1.6 clone could be picked up from AnTat 1.6 derived clones (N. Van M., unpublished data).

AnTat 1.6 is one of the major metacyclic VATs of the AnTAR I repertoire^{17,18}. It amounts to a fairly constant proportion (6–8%) of the metacyclic population, regardless of the VAT ingested by the tsetse fly¹⁹. The alterations of the VSA gene repertoire observed after AnTat 1.6 gene activation should thus occur frequently during cyclical transmission through the vector. In fact, we noticed that the cyclical transmission of an AnTat 1.1 expressor clone (AnTat 1.1D) through the fly leads to the loss of the AnTat 1.6 gene in the AnTat 1.1E clone (not shown); presumably the AnTat 1.1E clone was derived from an AnTat 1.6 metacyclic trypanosome.

We thank D. Franckx for preparation of the figures. This investigation received support from the Fonds de la Recherche Scientifique Médicale (FRSM, Brussels), the ILRAD/Belgian Research Centres Agreement for Collaborative Research (Nairobi) and the trypanosomiasis component of the UNDP/World Bank/WHO Special Programme for Research and Training in Tropical Diseases (Geneva).

Received 20 July; accepted 29 November 1983.

1. Hoeljmakers, J. H. J., Frasch, A. C. C., Bernards, A., Borst, P. & Cross, G. A. M. *Nature* **284**, 78–80 (1980).
2. Pays, E., Van Meirvenne, N., Le Ray, D. & Steinert, M. *Proc. natn. Acad. Sci. U.S.A.* **78**, 2673–2677 (1981).
3. Bernards, A. *et al. Cell* **27**, 497–505 (1981).
4. Pays, E. *et al. Cell* **34**, 371–381 (1983).
5. Young, J. R., Donelson, J. E., Majiwa, P. A. O., Shapiro, S. Z. & Williams, R. O. *Nucleic Acids Res.* **10**, 803–819 (1982).
6. Majiwa, P. A. O., Young, J. R., Englund, P. T., Shapiro, S. Z. & Williams, R. O. *Nature* **297**, 514–516 (1982).
7. Laurent, M. *et al. Nature* **302**, 263–266 (1983).
8. Pays, E. *et al. Cell* **34**, 357–369 (1983).
9. Borst, P. & Cross, G. A. M. *Cell* **29**, 291–303 (1982).
10. Williams, R. O., Young, J. R. & Majiwa, P. A. O. *Nature* **299**, 417–421 (1982).
11. Young, J. R., Shah, J. S., Matthyssens, G. & Williams, R. O. *Cell* **32**, 1149–1159 (1983).
12. Bernards, A., Michels, P. A. M., Lincke, C. R. & Borst, P. *Nature* **303**, 592–597 (1983).
13. Pays, E. *et al. Nucleic Acids Res.* **8**, 5965–5981 (1980).
14. Southern, E. M. *J. molec. Biol.* **98**, 503–517 (1975).
15. Weintraub, H. & Groudine, M. *Science* **193**, 848–856 (1976).
16. Pays, E. *et al. Cell* **35**, 721–731 (1983).
17. Le Ray, D., Barry, J. D., Easton, C. & Vickerman, K. *Ann. Soc. belge Méd. trop.* **57**, 369–381 (1977).
18. Le Ray, D., Barry, D. & Vickerman, K. *Nature* **273**, 300–302 (1978).
19. Hadjuk, S. L., Cameron, C. R., Barry, J. D. & Vickerman, K. *Parasitology* **83**, 595–607 (1981).
20. Van Meirvenne, N., Janssens, P. G. & Magnus, E. *Ann. Soc. belge Méd. trop.* **55**, 1–23 (1975).
21. Pays, E., Lheureux, M. & Steinert, M. *Nature* **292**, 265–267 (1981).

Genetically restricted antigen presentation for immunological tolerance and suppression

Adam Lowy, Jeffrey A. Drebin, John G. Monroe, Richard D. Granstein* & Mark I. Greene

Department of Pathology, Harvard Medical School, 25 Shattuck Street, Boston, Massachusetts 02115, USA

*Department of Dermatology, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts 02114, USA

The activation of some subsets of T cells requires the recognition of antigen in association with self Ia determinants^{1,2}. It is not clear, however, whether this is also necessary for the induction of unresponsiveness and the active suppression of hapten-specific T cells. We have studied the regulation of the delayed-type hypersensitivity (DTH) response in mice. Subcutaneous (s.c.) injection of haptenated syngeneic cells primarily activates T helper cells, while intravenous (i.v.) injection results in unresponsiveness and the activation of T suppressor cells³. Presentation of antigen by an I-A-positive antigen-presenting cell (APC) seems to be critical for T helper cell activation⁴. We now show that the activation of first-order T suppressor cells requires the presentation of hapten on an I-J-positive APC. However, i.v. injection of I-J-depleted haptenated cells also results in hapten-specific unresponsiveness. This non-transferable T-cell tolerance requires the presence of an I-A-positive APC and is genetically restricted. These results suggest that distinct modes of antigen presentation and administration are required for immunity, suppression and non-transferable tolerance.

We have previously demonstrated that presentation of antigen by a cyclophosphamide-sensitive I-J-positive APC is required for activation of third-order azobenzenearsonate (ABA)-induced T suppressor cells (Ts3)⁵. This APC may also be significant in the activation of T suppressor cells specific for other antigens, in systems where I-J-restricted antigen presentation has been demonstrated^{6–8}. Table 1a shows data indicating that the I-J APC is also essential for first-order T suppressor cell (Ts1) activation. Mice injected i.v. with ABA-coupled syngeneic spleen cells produce afferently acting Ts1 factor (TsF₁) which suppresses the DTH response in immune recipients. However, mice given ABA coupled to I-J-depleted cells, cannot produce ABA-specific transferable suppression. One might then predict

Table 1 Presentation of hapten by I-J APC is required for Ts1 induction but not for induction of non-transferable unresponsiveness

<i>a</i>	Subcutaneous injection (immunogen)	Factor generated by i.v. administration of	DTH	<i>P</i>
(1)	ABA-C3H	—	25.5 ± 1.3	—
(2)	ABA-C3H	ABA-C3H	10.3 ± 2.1	<0.005
(3)	ABA-C3H	(Anti-I-J ^k +C')ABA-C3H	33.0 ± 6.0	NS
(4)	—	—	11.5 ± 2.0	<0.05
<i>b</i>	Immunogen	Intravenous administration (tolerogen)		
(1)	ABA-C3H	—	19.2 ± 1.0	—
(2)	(Anti-I-J ^k +C') ABA-C3H	—	23.7 ± 1.7	NS
(3)	—	(Anti-I-J ^k +C')-ABA-C3H	9.5 ± 2.8	<0.05
(4)	—	—	8.9 ± 1.3	<0.05
<i>c</i>	Immunogen	Tolerogen		
(1)	ABA-C3H	—	15.5 ± 3.0	—
(2)	ABA-C3H	ABA-C3H	7.0 ± 0.9	<0.05
(3)	ABA-C3H	(Anti-I-J ^k +C')ABA-C3H	6.8 ± 0.9	<0.05
(4)	—	—	4.8 ± 1.4	<0.05

C3H/HeJ mice were immunized by s.c. injection of 3×10^7 ABA-coupled syngeneic spleen cells. In the experiment described in *a*, mice then received i.v. injection of factor on days 0–4. Factor was generated from spleen cell freeze-thaw extracts of mice which had received an i.v. injection of either 3×10^7 ABA-coupled syngeneic spleen cells (group 2) or 3×10^7 ABA-coupled syngeneic spleen cells wherein the cells had been treated with a monoclonal anti-I-J^k antibody (WF8.C12.8, given by Dr C. Waltenbaugh) and rabbit complement, C' (Pel Freeze; Rogers AR) (group 3). In *b* and *c*, mice received 3×10^7 ABA-coupled cells subcutaneously and/or intravenously on day 0. In groups 2 and 3 of experiment *b* and in group 3 of experiment *c*, spleen cells were treated with monoclonal anti-I-J^k antibody and C' before ABA coupling. In all experiments, mice were challenged with 30 µl of activated ABA solution in the left hind footpad on day 5; 24 h later, swelling was compared with the right unchallenged footpad. The data thus generated were analysed by Student's *t*-test. NS, not significant. The swelling induced in ABA immune mice has previously been shown¹⁶ to be due to a hapten-specific DTH response.

that without the ability to activate suppressor cells, i.v. administration of I-J-depleted APC would result in immunization of the animal, but this does not seem to be the case (Table 1*b*). ABA immune mice which receive i.v. injections of ABA coupled to I-J-negative cells, are rendered unresponsive to the hapten (Table 1*c*). Hence, while I-J-restricted presentation of antigen is required for T suppressor cell activation, in the absence of this I-J signal, non-transferable hapten-specific immunoregulatory events occur. This tolerance mechanism is resistant to treatment with low doses of cyclophosphamide, which is known to interfere with T suppressor cell activation (data not shown).

We investigated whether presentation of antigen in the context of self determinants was required for non-transferable tolerance. (C57BL/6 × A/J)_F₁ mice were immunized by s.c. administration of ABA-coupled A/J cells and simultaneously made tolerant by i.v. injection of I-J-depleted ABA-coupled C57BL/6 cells. As shown in Table 2, when the parental haplotypes of the priming and tolerizing cells differ throughout the entire H-2 region, animals are not rendered tolerant to the antigen. If I-J-positive cells are not depleted from the tolerizing cells, T suppressor cells are activated (Table 2, group 2). This implies that genetically restricted antigen presentation is

required for antigen-specific immunological unresponsiveness in the absence of T suppressor cell activation.

To study the genetic restrictions involved in tolerance induction in the absence of active suppression, we used APCs from cyclophosphamide-treated mice. We have shown previously that these mice are unable to donate APCs required for T suppressor cell induction⁵ although they can nonetheless be rendered tolerant in the presence of low doses of cyclophosphamide (data not shown). Our results (Table 3) demonstrate that (BALB/c × A/J)_F₁ mice immunized with X-ray-irradiated ABA-B10.A (*K^k I-A^k I-J^k I-E^k I-C^d D^d*) spleen cells could be rendered tolerant by i.v. administration of cells from cyclophosphamide-pretreated ABA-B10.A mice but not by similar administration of ABA-coupled cells from cyclophosphamide-pretreated B10.D2 (*K^d I-A^d I-J^d I-E^d I-C^d D^d*) mice. We further mapped the genetic restriction of tolerance induction by using B6 hapten-coupled cells to tolerize (B6A)_F₁ recipients. As shown in Table 4, I-A-depleted cell populations are not effective tolerogens for ABA-specific immune responses, and also cannot prime for ABA-specific immune responses. Hence, mice can be tolerized only by hapten coupled to cells which share I-A determinants with the cells used for immunization. We have already demonstrated that I-J-positive cells are not required

Table 2 Presentation of hapten for induction of non-transferable immunological tolerance is genetically restricted

Immunogen	Tolerogen	DTH	Probe
(1) ABA-A/J	—	23.5 ± 1.2	—
(2) ABA-A/J	ABA-C57BL/6	11.5 ± 2.5	<i>P</i> < 0.005
(3) ABA-A/J	(Anti-I-J ^b +C') ABA-C57BL/6	21.3 ± 1.2	NS
(4) —	—	5.8 ± 1.1	<i>P</i> < 0.001

The protocol is identical to that described in Table 1 with the following modifications: (C57BL/6 × A/J)_F₁ mice were used; mice received ABA-coupled, irradiated (1,500 R) A/J or C57BL/6 cells; and monoclonal anti-I-J^b antibody (WF 9.40.5, given by Dr C. Waltenbaugh), was used for lysis of I-J^b-positive C57BL/6 spleen cells. The reciprocal experiment (immunization with ABA-B6, tolerization with ABA-A/J) gave comparable results.

Table 3 K and/or I-A or E region-restricted antigen presentation for the induction of tolerance

Immunogen	Tolerogen	DTH	Probe
(1) ABA-B10.A	—	28.5 ± 2.5	—
(2) ABA-B10.A	Cyclophosphamide- pretreated ABA-B10.A	8.0 ± 0.9	<i>P</i> < 0.001
(3) ABA-B10.A	Cyclophosphamide- pretreated ABA-B10.D2	29.8 ± 2.3	NS
(4) —	—	8.5 ± 1.4	<i>P</i> < 0.001

The protocol is identical to that described in Table 1 with the following modifications: (BALB/c × A/J)_F₁ mice were injected with irradiated ABA-B10.A or ABA-B10.D2, and mice donating cells for tolerization were pretreated with cyclophosphamide (20 mg per kg intraperitoneal; Mead Johnson) 2 days before being killed.

Table 4 I-A-depleted cells are deficient in ability to tolerize T_{DTH} cells

	Immunogen	Tolerogen	DTH	P
(1)	ABA-C57BL6	—	36.5 ± 2.9	<0.005
(2)	ABA-C57BL6	(ABA-C57BL6)	6.8 ± 0.9	NS
(3)	ABA-C57BL6	(Anti-I-A ^b , anti-I-J ^b +C')	23.3 ± 2.3	<0.005
(4)	(Anti-I-A ^b , anti-I-J ^b +C') ABA-C57BL6	—	6.8 ± 1.8	NS
(5)	—	—	4.8 ± 1.5	—

The protocol is identical to that described in Table 2 with the following modifications: I-A, I-J-depleted cells were treated with anti-I-A^b antibody (25-9-17S, from the American Type Culture Collection, Rockville, Maryland) and anti-I-J^b antibody (WF 9.40.5) with complement sequentially.

for tolerance induction (Table 1c). Therefore, in order for hapten-specific T-cell tolerance to be induced, mice must be tolerized by hapten-coupled cells which share I-A determinants with the immunizing cells. We⁵ and others⁷⁻⁹ have previously described similar genetically restricted events in the activation of T suppressor and T helper cells. I-J-restricted presentation is required for ABA-specific T suppressor cell activation, while I-A-restricted presentation is needed for T helper cell activation. These results demonstrate that activation of non-transferable immune tolerance also requires genetically restricted antigen presentation.

This work suggests different mechanisms for the induction of T- and B-cell tolerance. As B-cell receptors are able to perceive nominal antigen, not modified by self determinants, unresponsiveness may be induced by free antigen¹⁰. In contrast, the I-A-restricted T helper cell receptor may require presentation of antigen in the context of self determinants for the induction of unresponsiveness¹¹. This would be consistent with earlier work suggesting a difference between the requirements for T- and B-cell tolerance¹² and with the suggestion that T-cell tolerance to minor histocompatibility antigens may require antigen processing by an accessory cell¹³. Previous studies of T-cell tolerance to specific hapten have indicated that hapten coupled to cell membranes increases the efficiency of the tolerogenic signal^{14,15}.

One might postulate from our data that T helper cells are only susceptible to the tolerogenic signal when they are activated to hapten plus syngeneic I-A. Presentation of antigen in the context of self I-A determinants is then required for these T helper cells to be rendered unresponsive. A possible explanation for this mechanism would be that the receptors of effector T_{DTH} cells, which we have previously shown to be I-A-restricted in their activation¹⁶, are blocked by intravenously administered haptenated cells.

From these studies on the hapten-specific Lyt 1⁺-dependent DTH response to ABA, we have been able to demonstrate that distinct modes of antigen presentation and administration are required for immunity (s.c. administration of hapten plus I-A), suppression (hapten plus I-J) and non-transferable tolerance (i.v. administration of hapten plus I-A). It may be that antigen presentation by cells expressing different Ia antigens affects the immune response by leading to autoimmunity or immunodeficiency. Hence, manipulation of differential antigen presentation may have clinical consequences. In fact, administration of monoclonal anti-I-A antibodies to autoimmune mice decreases autoreactive events in these animals¹⁷. This result would be expected if anti-I-A antibodies blocked presentation of self antigen to T helper cells and led to presentation to T suppressor cells by the I-J⁺ APC. Such administration of anti-I-A antibodies does convert immunity to suppression^{18,19}. Furthermore, preliminary evidence from our laboratory (A.L. and M.I.G., manuscript in preparation) suggests that a primary effect

of *in vivo* anti-I-A administration may be at the APC level. As we are able to distinguish between antigen presentation mechanisms for tolerance or suppression, we may be able to investigate the role of these mechanisms in the establishment of self tolerance and autoimmunity.

This work was supported by grant AI-19901-01. J.A.D. was supported by grant CA-09141 from the NIH in conjunction with the MD, PhD Program at Harvard Medical School; J.G.M. was the recipient of an Aid to Cancer Research Postdoctoral Fellowship; R.D.G. was supported by NRSA IF32CA07014-02; and M.I.G. was the recipient of a faculty research award from the American Cancer Society. We thank Drs R. L. Erikson, M. Mescher and J. S. Bromberg for helpful discussions; and Ms Sarah Curwood for secretarial assistance.

Received 26 October 1983; accepted 31 January 1984.

- Shevach, E. M. & Rosenthal, A. S. *J. exp. Med.* **138**, 1213-1229 (1973).
- Schwartz, R. H., David, C. S., Dorf, M. E., Benacerraf, B. & Paul, W. *Proc. natn. Acad. Sci. U.S.A.* **75**, 2387-2391 (1978).
- Chase, M. W. *Proc. Soc. exp. Biol. Med.* **61**, 257-259 (1946).
- Sprent, J., Korngold, R. & Molnar-Kimbar, K. *Springer Semin. Immunopath.* **3**, 213-245 (1980).
- Lowy, A., Tominaga, A., Drebin, J., Benacerraf, B. & Greene, M. I. *J. exp. Med.* **157**, 353-358 (1983).
- Gershon, R. K. *Contemp. Topics Immunobiol.* **3**, 1-40 (1974).
- Zembala, M., Asherson, G. & Colizzi, V. *Nature* **297**, 411-413 (1982).
- Minami, M., Nonji, N. & Dorf, M. E. *J. exp. Med.* **156**, 1502-1515 (1982).
- Green, E. F. & Colley, D. G. *Proc. natn. Acad. Sci. U.S.A.* **78**, 1152-1156 (1981).
- Nossal, G. & Schrader, J. W. *Transplant. Rev.* **23**, 138-158 (1975).
- Claman, H. *Ann. N. Y. Acad. Sci.* **393**, 291-299 (1982).
- Chiller, J. M., Habicht, G. S. & Weigle, W. O. *Science* **171**, 813-815 (1971).
- Matzinger, P. & Waterfield, J. D. *Nature* **285**, 492-494 (1980).
- Battisto, J. & Bloom, B. *Nature* **212**, 156-157 (1966).
- Claman, H. N., Miller, S. D., Sy, M.-S. & Moorehead, J. W. *Immun. Rev.* **50**, 105-132 (1980).
- Bach, B. A., Sherman, L., Benacerraf, B. & Greene, M. I. *J. Immun.* **121**, 1460-1468 (1978).
- Waldor, M. K., Sriram, S., McDevitt, H. O. & Steinman, L. *Proc. natn. Acad. Sci. U.S.A.* **80**, 2713-2717 (1983).
- Berzofsky, J. A. & Richman, L. K. *J. Immun.* **126**, 1898-1904 (1981).
- Perry, L. L. & Greene, M. I. *J. exp. Med.* **156**, 480-491 (1982).

Cofactor requirements of splicing of purified messenger RNA precursors

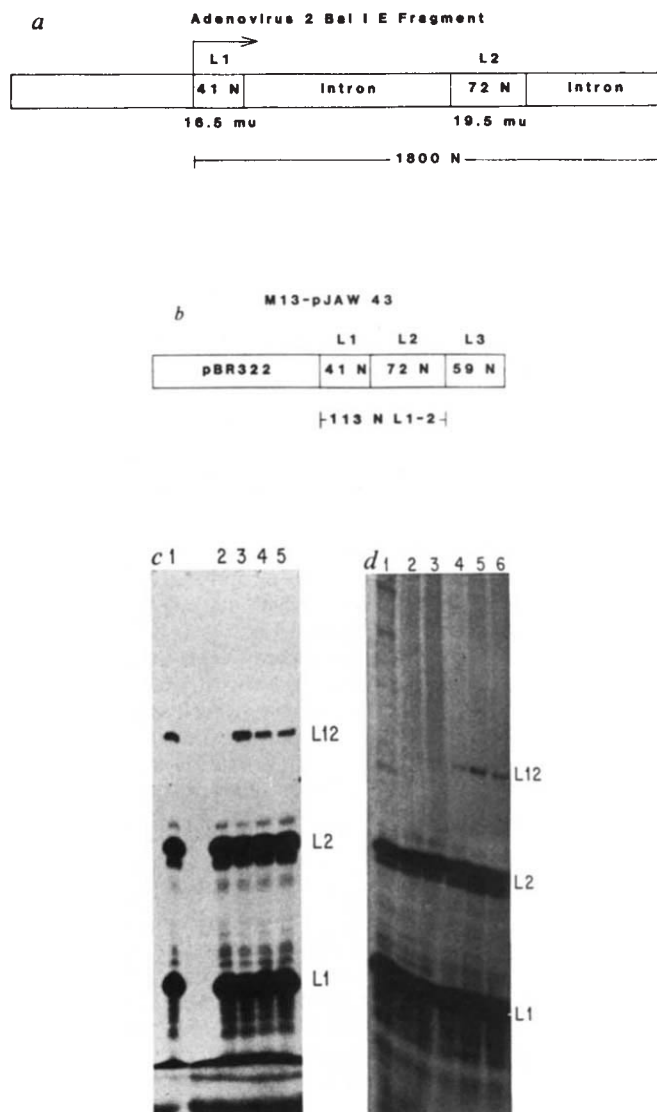
Stephen F. Hardy, Paula J. Grabowski, Richard A. Padgett & Phillip A. Sharp

Center for Cancer Research and Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

The origin and functions of introns in protein coding genes is one of the enigmas of molecular biology. Splicing processes that remove intervening sequences from precursor RNAs must have either predated or co-evolved with introns. Inferences about the origin of introns and the possible modes of regulation of splicing should emerge from an understanding of the biochemical mechanisms of splicing. The biochemistry of splicing of tRNA^{1,2} and rRNA^{3,4} precursors has rapidly advanced with the development of *in vitro* reactions containing soluble components that duplicate *in vivo* reactions. We have recently shown that accurate splicing of an adenovirus mRNA precursor occurs during a coupled transcription/splicing reaction in a soluble whole cell extract⁵. We now report that an exogenous RNA substrate containing the first and second leaders of adenovirus 2 is accurately spliced when added to an extract of HeLa cells. ATP and Mg²⁺ are essential cofactors for the reaction. The time course of splicing is unusual; a lag of 45 min is observed before the appearance of spliced product.

Exogenous RNA substrate (1,800 nucleotides) was synthesized *in vitro* in the whole cell extract in conditions in which splicing is inhibited (see Fig. 1 legend for method). Subsequently the purified ³²P-labelled RNA was incubated in the whole cell extract in conditions which allow splicing. The RNA products of the reaction were analysed by a sensitive technique, described previously⁵, involving hybridization to the single-stranded M13

Fig. 1 A purified exogenous substrate RNA is spliced in a whole cell extract. **a**, Structure of the adenovirus 2 (Ad2) *BalI* fragment E DNA template used for the synthesis of RNA precursor. (N, number of nucleotides.) The *BalI* E fragment contains 679 bp of 5' noncoding sequences, the major late promoter, the L1 (41N) exon, a 1,020-base pair (bp) intron, the L2 (72N) exon and 578 bp of the second intron. **b**, Structure of the M13 cDNA recombinant used to analyse the products of *in vitro* splicing. The *EcoRI*-*XhoI* fragment of pJAW 43 (ref. 13) was inserted between the *EcoRI* and *SalI* sites of M13 mp9. This fragment contained the *PstI*-*EcoRI* fragment of pBR322 and the first and second and part of the third leader (L1-2-3) of Ad2. **c**, Determination of nucleoside triphosphate requirement for splicing. Gel-purified RNA precursor was incubated in splicing conditions in the presence of three of the four nucleoside triphosphates. Each splicing reaction contained 50,000 c.p.m. RNA precursor per reaction, 28 mM KCl, 3.5 mM MgCl₂, 5 mM creatine phosphate, 4.8% glycerol, 5.6 mM HEPES, pH 7.6, 0.1 mM EDTA, 0.6 mM dithiothreitol and 7 μ l extract in a total volume of 25 μ l. Incubations were carried out at 30 °C for 3 h. The RNA products of the reaction were hybridized to the M13 DNA shown in **b** followed by digestion with pancreatic RNase. The protected RNA was recovered from the hybrid and analysed by electrophoresis on an 8% polyacrylamide/8 M urea gel. Lane 1, a control in which all four nucleoside triphosphates, ATP, GTP, CTP and UTP, were present. Lanes 2-5 were reactions incubated in the same conditions, but without ATP (lane 2), GTP (lane 3), CTP (lane 4) and UTP (lane 5). Each nucleoside triphosphate was present in the reactions at a final concentration of 1 mM. **d**, Extract titration. Varying amounts of whole cell extract were added to splicing reactions at a final volume of 25 μ l, followed by analysis of the RNA products as described in **c**. Lanes: 1, 0.5 μ l; 2, 1.0 μ l; 3, 2.0 μ l; 4, 4.0 μ l; 5, 6.0 μ l; and 6, 8.0 μ l extract. RNA precursor was produced by transcription *in vitro* in a whole cell extract in the following conditions: 50% extract prepared as described previously¹⁴, 50 mM KCl, 10 mM HEPES pH 7.9, 6.25 mM MgCl₂, 0.25 mM EDTA, 1 mM dithiothreitol, 100 μ M of ATP, CTP and GTP, 5 μ M UTP including 1 mCi [α -³²P]UTP, 5 mM creatine phosphate and 25 μ g ml⁻¹ DNA template (panel **a**). Transcription was carried out at 30 °C for 2 h. The reaction was stopped by the addition of one volume of 300 mM NaCl, 100 mM Tris pH 7.9, 10 mM EDTA, 2% SDS and proteinase K to a final concentration of 250 μ g ml⁻¹, followed by an additional incubation of 15 min at 30 °C. The 1,800-nucleotide (N) RNA precursor was purified by elution from a 3.5% polyacrylamide/8 M urea gel for the experiment of **c** or by Sephadex G50-150 column chromatography for the experiment of **d**. The protein concentration of the extract used for the experiment of **d** was ~20 mg ml⁻¹.



cDNA shown in Fig. 1b followed by RNase treatment and binding to nitrocellulose filters. Digestion of the RNA/DNA hybrids with pancreatic RNase yields three possible RNAs, L1 (41 nucleotides) and L2 (72 nucleotides) leaders from unspliced substrate and L12 (113 nucleotides) spliced product.

Both optimal conditions and cofactor requirements were determined for splicing of the exogenous RNA substrate. The critical components in the previously characterized coupled system were ATP, GTP, CTP, UTP, KCl, Mg²⁺ and creatine phosphate. Figure 1c shows an experiment in which each of the four nucleoside triphosphates was omitted from the reaction. The result suggests that ATP is an essential cofactor for splicing (lane 2c) and that none of the other nucleoside triphosphates can substitute for the ATP requirement. In the presence of creatine phosphate (5 mM), the percentage of L12 RNA increased linearly with increasing ATP concentration (Fig. 2b). Titration of ATP in the absence of creatine phosphate resulted in no splicing activity at ATP concentrations below 5 mM. This is probably due to the instability of ATP in a whole cell extract. These extracts contain creatine phosphokinase which in the presence of creatine phosphate recharges nucleoside triphosphates. Since the whole cell extract contains a free pool of approximately 1 μ M nucleotides⁶, we cannot rule out the possibility that a very small amount of a nucleotide other than ATP is required.

Figure 2 shows titrations of Mg²⁺ and monovalent cation in the presence of 1 mM ATP. The optimal Mg²⁺ concentration was approximately 1 mM Mg²⁺, while 5 mM Mg²⁺ totally inhibited the reaction. The optimal concentration of K⁺ was

found to be 28 mM, while higher or lower concentrations resulted in a linear decrease in splicing activity. Since only a small fraction of the RNA precursor was spliced in a typical reaction, it was interesting to test whether the level of substrate RNA saturated the reaction. This is unlikely since a constant percentage of the precursor was converted to spliced RNA, when the amount of added RNA precursor was varied over a 125-fold range (data not shown). Figure 1d shows an autoradiogram of a gel resolving the products of an extract titration. The reaction exhibited a nonlinear dependence on protein concentration with an optimum of about 5 mg extract protein ml⁻¹. This optimum may vary with different extracts as has been observed for transcription activity in whole cell extract⁷.

In optimal conditions for the coupled transcription/splicing reaction, only 4-5% of the precursor RNA was converted to spliced product. Titration of salt and cofactors in the uncoupled reaction has resulted in an increase in spliced RNA product. In the experiment shown in Fig. 2c, 13% conversion to spliced RNA was obtained. A summary of the titration results suggests that optimal conditions are: 28 mM KCl, 2.5 mM Mg²⁺, 2 mM ATP, 5 mM creatine phosphate and 5 mg extract protein ml⁻¹.

We have previously reported that 3-h incubations were necessary for the synthesis of spliced RNA in a coupled reaction⁵. A similar time course for splicing was observed with exogenously added RNA precursor (Fig. 2a). No detectable spliced RNA was produced during the first 45 min of incubation. The reaction rate increased markedly after a 60-min lag, then continued at this rate for an additional 30-60 min. The maximum amount of spliced RNA was obtained after 180 min

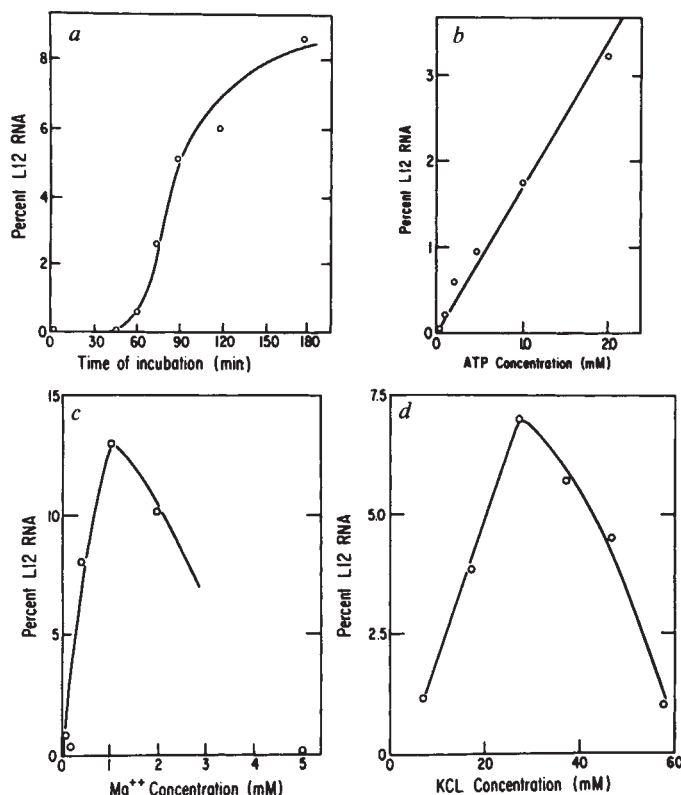


Fig. 2 Time course and cofactor requirements for splicing *in vitro* of an exogenous RNA precursor. *a*, Purified precursor RNA was incubated for various times in the whole cell extract in splicing conditions. The RNA products of the reaction were then analysed by hybridization to M13 pJAW 43 DNA followed by digestion with pancreatic RNase, as described in Fig. 1. The fraction of spliced RNA was determined by densitometric tracing of the autoradiogram. The ATP dependence, *b*, the Mg²⁺ dependence, *c*, and the KCl dependence, *d*, were determined in analogous experiments. RNA precursor synthesized as described in Fig. 1 was purified by Sephadex G-50-150 chromatography to remove the nucleoside triphosphates. The splicing reactions were performed as described in Fig. 1, except for the following modifications. For the experiment shown in *a*, each reaction contained a concentration of Mg²⁺ equal to the ATP concentration plus 0.5 mM. For the experiment of *c*, ATP was present at a concentration of 1 mM. For the experiment of *d*, ATP and Mg²⁺ were present at a concentration of 1 mM and 1.5 mM, respectively.

of incubation. This level of spliced product remained stable for an additional 12-h incubation. The unusually long lag cannot reflect a slow ATP-independent step involving assembly of complexes which then rapidly utilize ATP to cleave and ligate RNA, because preincubation of labelled RNA and extract in the absence of ATP does not shorten the lag period between the addition of ATP and the appearance of spliced product. This suggests that the rate-limiting step, perhaps formation of a complex with substrate RNA, is ATP dependent.

Cofactor requirements for the splicing of mRNA precursors in an extract of HeLa cells are not similar to those of rRNA splicing which does not require ATP^{3,4} but are similar to those for the splicing of tRNA precursors; both reactions require ATP and Mg²⁺ (refs 8, 9). The two reactions differ in their Mg²⁺ dependence, however; 10 mM Mg²⁺ is optimal for tRNA splicing² while mRNA precursor splicing is inhibited at concentrations of Mg²⁺ above 5 mM. In addition, whereas other nucleoside triphosphates such as UTP, CTP and GTP are able to substitute for the ATP requirement in the tRNA splicing reaction², these nucleoside triphosphates are not able to substitute for the ATP requirement in the mRNA splicing reaction (Fig. 1c). These differences may reflect the use of different enzymatic activities by the two types of RNA precursors.

The splicing of tRNA precursors occurs by two distinct and separable steps: endonucleolytic cleavage followed by ligation

of the exons^{8,9}. Whereas the only cofactor required for the cleavage reaction is Mg²⁺, both Mg²⁺ and ATP are necessary for the ligation reaction. It is unlikely that a similar division of the mRNA splicing reaction is possible. At present, we have not found that a significant fraction of the exogenously added mRNA precursor is cleaved when incubated in a reaction containing Mg²⁺ but no ATP (S.F.H. and P.J.G., unpublished observations). Also, preincubation of RNA and extract in the absence of ATP does not result in the formation of complexes that quickly process RNA upon addition of the cofactor.

We have previously shown that addition of antisera that react with the U1 small nuclear ribonucleoprotein particles (U1 RNP) inhibited splicing in the coupled transcription/splicing system¹⁰. Similar experiments using exogenous substrate RNA also show inhibition of splicing by addition of the same anti-U1 RNP sera. The pronounced lag before appearance of spliced product might reflect the slow assembly of an RNP-substrate complex in the reaction mix. Formation of an RNA/RNA duplex by sequences at the 5' end of U1 RNA and the 5' splice site has been suggested as an important step in recognition of the mRNA precursor^{11,12}. However, there is no inherent reason why this type of recognition should be slow given that the U1 RNA concentration in the reaction is 5 µg ml⁻¹.

P.J.G. acknowledges support from the Helen Hay Whitney Foundation and R.A.P. from the Myron A. Bantrell Charitable Trust. This work was supported by NSF grant PCM-7823230 (currently PCM-8200309), NIH grant P01-CA26717 to P.A.S. and partially from Center for Cancer Biology at MIT (Core) grant NIH-P01-CA14051.

Received 21 October 1983; accepted 6 January 1984.

1. Peebles, C. L., Gegenheimer, P. & Abelson, J. *Cell* **32**, 525-536 (1983).
2. Greer, C. L., Peebles, C. L., Gegenheimer, P. & Abelson, J. *Cell* **32**, 537-546 (1983).
3. Cech, T. R., Zaugg, A. J. & Grabowski, P. J. *Cell* **27**, 487-496 (1981).
4. Kruger, K. *et al.* *Cell* **31**, 147-157 (1982).
5. Padgett, R. A., Hardy, S. F. & Sharp, P. A. *Proc. natn. Acad. Sci. U.S.A.* **80**, 5230-5234 (1983).
6. Fire, A., Baker, C. C., Manley, J. L., Ziff, E. B. & Sharp, P. A. *J. Virol.* **40**, 703 (1981).
7. Manley, J. L., Fire, A., Samuels, M. & Sharp, P. A. *Meth. Enzym.* **101**, 568-582 (1983).
8. Peebles, C. L., Ogden, R. C., Knapp, G. & Abelson, J. *Cell* **18**, 27-35 (1979).
9. Knapp, G., Ogden, R. C., Peebles, C. L. & Abelson, J. *Cell* **18**, 37-45 (1979).
10. Padgett, R. A., Mount, S. M., Steitz, J. A. & Sharp, P. A. *Cell* **35**, 101-107 (1983).
11. Lerner, M. R., Boyle, J. A., Mount, S. M., Wolin, S. L. & Steitz, J. A. *Nature* **283**, 220-224 (1980).
12. Rogers, J. & Wall, R. *Proc. natn. Acad. Sci. U.S.A.* **77**, 1877-1879 (1980).
13. Zain, S. *et al.* *Cell* **16**, 851-861 (1979).
14. Manley, J. L., Fire, A., Cano, A., Sharp, P. A. & Gelfand, M. L. *Proc. natn. Acad. Sci. U.S.A.* **77**, 3855-3859 (1980).

Similarity of the nucleotide sequences of rat α -lactalbumin and chicken lysozyme genes

Pradman K. Qasba & Surinder K. Safaya

Laboratory of Pathophysiology, National Cancer Institute, Bethesda, Maryland 20205, USA

α -Lactalbumin (α -LA) is a milk protein that interacts with the enzyme galactosyltransferase, modifying its substrate specificity in a way which promotes the transfer of galactose to glucose, resulting in a β -1 $\rightarrow$ 4 glycosidic linkage and the synthesis of lactose^{1,2}. Lysozyme, an enzyme which catalyses the hydrolysis of a β -1 $\rightarrow$ 4 glycosidic linkage in polysaccharides, has been shown to be structurally related to α -LA and it has been proposed that they have arisen from a common ancestral gene³. To compare their evolutionary relationships, we report here the complete nucleotide sequence of the rat α -LA gene, including its 5'-flanking sequences, and compare its gene structure with the chicken egg-white lysozyme gene⁴. Both genes contain three introns at similar positions. The first three exons of the two genes have similar nucleotide sequences. The fourth exon of α -LA, which partly codes for the C-terminal residues of the protein, essential for its interaction with galactosyltransferase^{5,6}, is markedly different from the corresponding exon of the lysozyme gene and is preceded by two (TG)_n repeats.

Fig. 1 (left) Sequence of the sense strand of the rat α -LA gene region. The coding sequences of the exons have been translated and the amino acid residues numbered below. The start of RNA transcription was identified at position 1,247 (▼) by primer extension¹⁰ and at position 1,245 (▽) by S_1 nuclease mapping¹¹. The termination site of the RNA transcript at position 3,762 is shown by †. The consensus sequence, TAAATAAAA, and the polyadenylation site are boxed. The short repeat sequences in the 5'-flanking region and unique sequences of the intervening region are underlined.

Methods: A 32 P-labelled rat α -LA cDNA clone p α -LA 18⁷ was used as a hybridization probe to screen 5×10^6 plaques from the bacteriophage Charon 4A/rat partial *Eco*RI genomic library⁸. Seven individual hybridizing plaques were isolated, plaque purified, and DNA from each recombinant phage was isolated and mapped with different restriction endonucleases. Rat α -LA cDNA has a single *Eco*RI site⁷. The 3' and 5' end fragments to this site were used as hybridization probes. Comparison of the hybridization patterns of the phage DNAs and the restriction enzyme maps showed that there were two classes of recombinant phages. One class contains 6 and 2.7 kb *Eco*RI liver DNA fragments. Only the 6 kb fragment hybridized to the 3' end probe of p α -LA cDNA. The other class contains *Eco*RI fragments of approximately 0.9, 2.5, 1.0, 5.5, 1.1 and 0.5 kb. Only 0.9 and 2.5 kb fragments hybridized to the 5'-end probe of the essentially full-length α -LA cDNA clone, p α -LA 35 (short of 4 nucleotides on the 5' end) or to α -LA mRNA labelled at the 5' end. A recombinant phage carrying overlapping α -LA genomic DNA fragment of the two phage classes could not be identified from this partial *Eco*RI genomic library even after additional screening of 2×10^6 phages. The restriction DNA fragment of the α -LA gene region from rat mammary gland or liver DNA could be arranged into a genomic map indistinguishable from the restriction enzyme map derived from the two recombinant phages. We have recently isolated a recombinant phage carrying an overlapping α -LA rat genomic DNA from a bacteriophage Charon 4A/rat partial *Hae*III genomic library. Rat *Eco*RI DNA fragments from each phage DNA were isolated on 1% low melting agarose gel and subcloned into the *Eco*RI site of pBR322. Plasmid DNA and their *Eco*RI inserts were isolated and purified from such subclones. The linkage of the *Eco*RI fragments of the recombinant phage DNAs was determined by partial digestion of the DNA with *Eco*RI or extensive digestion with a combination of restriction enzymes, followed by Southern blotting²⁹ and hybridization with the 32 P-labelled *Eco*RI inserts of the subclones. The inserts of some of the subclones, p α -LA 2.5 and p α -LA 0.9, containing 2.5 and 0.9 kb *Eco*RI liver DNA fragments, respectively, were extensively mapped with restriction endonucleases and sequenced by the Maxam and Gilbert method³⁰. The nucleotide sequence from positions 1–2,574 and 2,575–3,522 was determined from p α -LA 2.5 and p α -LA 0.9, respectively. The remaining sequence from position 3,523–3,829 was determined from the *Eco*RI/*Bam*HI fragment isolated from the 6 kb genomic rat DNA fragment.

The nucleotide sequence of the entire rat α -LA gene and its 5' flanking region is shown in Fig. 1. The α -LA gene extends over 2.5 (kb kilobases) of genomic DNA and is composed of four exons and three introns. The nucleotide sequence of the coding region of the genomic DNA is consistent with the previously reported cDNA sequences of p α -LA 18 and p α -LA 17⁷ and with the essentially full-length α -LA cDNA clone p α -LA 35 (Fig. 1 legend). The amino acid sequence of the entire rat α -LA protein has been deduced from the DNA sequence of the gene. This has confirmed our previously published partial sequence of rat α -LA⁷ which showed that rat α -LA has a 17 residue extension at the carboxy terminus which is proline rich and hydrophobic. The nucleotide sequence of the gene shows

that this extension results from a T to G base change in the termination codon. Comparison of the nucleotide deduced amino acid sequence with the amino acid sequence of rat α -LA reported by Prasad *et al.*⁹ shows certain differences. We believe that some of these discrepancies are probably a result of an incorrect amino acid assignment made because of deamination of some amino acids during the protein sequence analysis. We cannot explain the remaining discrepancies.

The 5' end of the rat α -LA mRNA was located at G, position 1,247, by the primer extension method¹⁰. However, it was identified as T at position 1,245 (Fig. 1) by S_1 nuclease mapping¹¹. Preceding this initiation site is an A-rich sequence, 5'-TAAATAAAA-3', at position 1,224–1,231. This is similar but not identical to the consensus sequence TATAAA¹² found at the initiation site of many other genes. There are also a number of almost identical short repeat sequences preceding this A-rich sequence, shown underlined in Fig. 1. These appear only in the 5' flanking region of this gene. It remains to be established whether these short repeat sequences are in any way related to the induction or inhibition of the α -LA gene expression by multiple hormonal combinations¹³. It is worth noting that a nonanucleotide sequence ATCCCTTTC, which appears at positions 748–756 and 860–868, resembles the first 9 nucleotides of the 19-nucleotide consensus sequence ATCCCTTTCATCTGTTTGTGTA thought to be involved in the progesterone receptor recognition site in the ovalbumin gene¹⁴. The three intervening sequences of the α -LA gene have some unusual features (Fig. 1). In the first intervening sequence a trinucleotide, TCC, is repeated 23 times between positions 1,459 and 1,528. In the third intervening sequence TG dinucleotide is repeated 25 times from position 2,897 to 2,946 and again 22 times from position 3,263 to 3,306. Between these two stretches of TG repeat sequences a trinucleotide, TAT, is repeated from position 2,964 to 3,114. The TG repeats have been shown to occur in other eukaryotic genomes^{15,16}.

Comparison of rat α -LA and chicken egg white lysozyme gene organization⁴ reveals that the three introns are located at similar positions within the coding regions of the two genes (Fig. 2). Rat α -LA exons are 165 (167), 159, 76 and 328 base pairs (bp) long, compared with the corresponding lysozyme exons which are 165, 162, 79 and 180 bp long. The three introns are located on codons 26, 79 and 104 within the coding regions of rat α -LA, compared with codons 28, 82 and 108 of chicken egg white lysozyme. The first and third introns in both the genes interrupt the Trp codon. However the corresponding intervening sequences in the two genes differ in length: in the rat α -LA gene the intervening sequences are 341, 429 and 1,016 bp long and in the chicken egg white lysozyme gene 1,270, 1,810 and 79 bp long. All three introns in both the genes not only begin with GT and end with an AG dinucleotide, sequences thought to be necessary for correct RNA splicing of various other eukaryotic genes¹⁷, but also have additional common sequences at the exon-intron junctions (Fig. 2).

Comparison of the nucleotide sequences of the coding regions of the first three exons of rat α -LA with chicken egg white lysozyme show sequence similarities of 43%, 56% and 46% in exons 1, 2 and 3, respectively. For this comparison, the amino acid sequences of the two proteins in these regions were first aligned and subsequently their nucleotide sequences. This required one additional residue (in the signal peptide) and two

Fig. 2 Comparisons of the rat α -LA and the chicken egg white lysozyme gene regions. Solid boxes represent the regions coding for the structural sequences of the protein; open boxes with a dashed line at the 5'- and at 3'-ends represent, respectively, the region coding the signal peptide of the protein and the 3' noncoding region. In exon 1 the dashed line represents the 5' noncoding region. Numbers represent the bp and Int the intervening sequences.

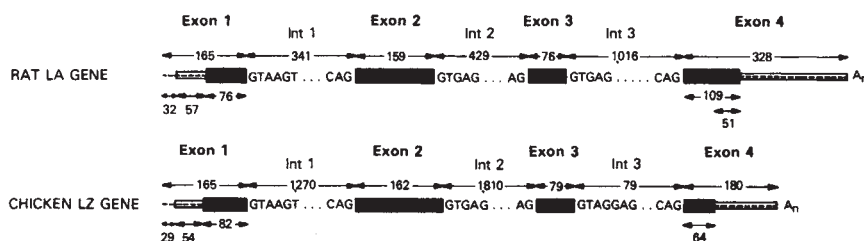
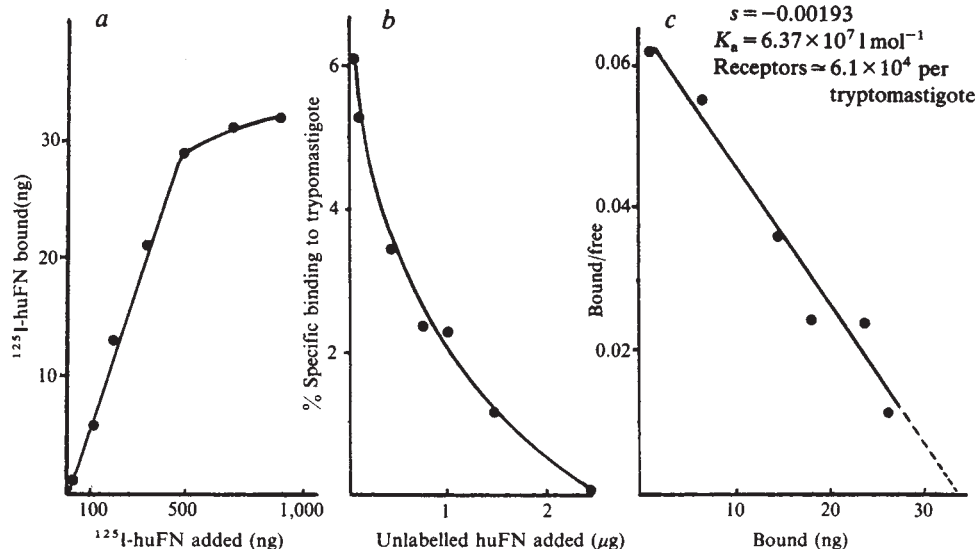


Fig. 1 Binding of ^{125}I -labelled human fibronectin (huFN) to glutaraldehyde-fixed *T. cruzi* trypomastigote culture forms (●). *a*, ^{125}I -huFN binding (in ng) to *T. cruzi* trypomastigote culture forms as a function of ^{125}I -ligand input. *b*, Binding of ^{125}I -huFN to trypomastigotes in the presence of increasing amounts of unlabelled ligand. The binding assay was done in a total volume of 250 μl . *c*, Scatchard¹⁷ analysis of the binding of ^{125}I -huFN to trypomastigotes. The fibronectin molecule consists of two similar or identical subunits of molecular weight (MW) $220,000 \pm 20,000$ (ref. 9). Calculations of the number of receptors and K_d were based on a MW of 220,000 and on the assumption that both monomeric subunits bind to fibronectin receptors. Each point represents the mean of three experiments.



Methods: Fibronectin was prepared according to the method of Ruoslahti *et al.*¹⁸. Citrated human blood was passed through Sepharose 4B-gelatin columns (Pharmacia) equilibrated with phosphate-buffered saline (PBS), 0.01 M citrate buffer. Fibronectin was eluted with 0.05 M Tris-HCl pH 7.5, containing 4 M urea. The product was passed through Sephadex G25 columns, equilibrated with 20 mM phosphate buffer pH 6.8, containing 1 mM EDTA and 100 mM NaCl to eliminate urea. Protein contaminants of the fibronectin solution were eliminated by passage through a DE 52 column (Whatman). Fibronectin was eluted by using a NaCl gradient (100–400 mM) in the buffer described above. The purity of the fibronectin solution was tested by polyacrylamide gel electrophoresis, then it was radiolabelled with ^{125}I to a specific activity of 2×10^6 c.p.m. μg^{-1} using the chloramine-T method¹⁹. The Tehuantepec strain of *T. cruzi* was used throughout²⁰. Trypomastigote culture forms were obtained from monolayers of 3T3 FR fibroblasts as described previously¹⁶; 1.5×10^6 trypomastigotes in 100 μl of Eagle's MEM were incubated for 1 h at 37 °C with ^{125}I -huFN in a total volume of 150 μl , in polystyrene tubes previously saturated with PBS containing 2% BSA to prevent binding of fibronectin to the tube surfaces. Next, 50 μl of the mixture were applied to replicate to 200 μl of dibutyl phthalate (Aldrich Chemical Co. Inc. Milwaukee, Wisconsin) in a 300- μl Eppendorf tube and centrifuged in an Eppendorf microfuge for 30 min at room temperature. The pellets, obtained by cutting the tips of the tubes with a razor blade, as well as the supernatants pooled from replicate tubes, were directly assayed for radioactivity in a γ -scintillation spectrometer. Nonspecific binding was determined in the presence of a 100-fold excess of unlabelled ligand and represented 14–17% of total binding. The per cent ^{125}I -huFN specifically bound was calculated by subtracting from the total binding the nonspecific binding measured as described above.

the mammalian cell surface 'receptors' which interact with the parasite is limited. We now report that fibronectin, which is a high molecular weight glycoprotein present in blood, connective tissue and at cell surfaces⁹, binds specifically to *Trypanosoma cruzi* trypomastigotes. The reaction is specific, reversible (in the presence of a 100-fold molar excess of unlabelled ligand) and of moderate affinity ($K_d = 11.36$ nM). Various other proteins (for example, thyroglobulin, ferritin, catalase, aldolase, human IgG and bovine serum albumin) had no significant effect on the binding of labelled ligand to the parasite surface. Addition of anti-fibronectin antibodies to the culture medium significantly inhibited the infection of rat fibroblasts (3T3 FR) by *T. cruzi* trypomastigotes, suggesting that cell surface fibronectin may act as a recognition site for attachment of the parasites.

Recently, fibronectins have elicited considerable interest⁹. The attachment of some bacteria to fibronectin on the host cell surface¹⁰ has been described, but its role in the cell-protozoa interaction has not yet been considered. Incubation of *T. cruzi* trypomastigotes with increasing amounts of ^{125}I -labelled human fibronectin (^{125}I -huFN) showed that the binding is a saturable process (Fig. 1*a*); about 33 ng of radioligand bound per 1.5×10^6 trypomastigotes was sufficient to achieve apparent saturation. Half-maximal saturation was reached with 250 ng of radioligand, providing an estimate of the equilibrium dissociation constant, K_d (11.36 nM), of the radioligand for specific binding sites on trypomastigotes. When trace amounts (25 ng) of the radiolabelled ligand were incubated with *T. cruzi* trypomastigotes in the presence of increasing amounts of unlabelled huFN, the latter competed efficiently with the radiolabelled ligand for binding to the parasite surface (Fig. 1*b*). From the concentration of unlabelled ligand at which the binding was reduced by 50%, a dissociation constant of 14.39 nM was calculated; this agrees with the value obtained using the radiolabelled ligand. There are about 6×10^4 fibronectin receptors per trypomastigote (Fig. 1*c*), as determined by Scatchard analysis of several experiments.

Table 1 Inhibition of binding of ^{125}I -huFN to *T. cruzi* trypomastigote culture forms

	% Inhibition ($\pm$ s.e.m.)
Unlabelled huFN	67 ± 12
Human IgG	5 ± 3
Thyroglobulin	7 ± 2
Ferritin	3 ± 1
Catalase	3 ± 2
Aldolase	4 ± 1
Bovine serum albumin	2 ± 1

1.5×10^6 trypomastigotes in 100 μl of MEM were incubated with 25 ng of ^{125}I -huFN (50 μl) without or with 1.5 μg of the following proteins: unlabelled human fibronectin, human IgG, thyroglobulin, ferritin, catalase, aldolase or BSA (Sigma) in a total volume of 250 μl . The binding assay was as described in Fig. 1 legend. Per cent inhibition was calculated as follows:

$$1 - \frac{\Delta \text{ c.p.m. (with inhibitor)}}{\Delta \text{ c.p.m. (without inhibitor)}} \times 100$$

where Δ c.p.m. (with inhibitor) is the radioactivity bound to trypomastigotes in the presence of inhibitor minus the radioactivity bound in the presence of a 100-fold excess of unlabelled huFN, and Δ c.p.m. (without inhibitor) is the radioactivity bound to trypomastigotes without inhibitor minus the radioactivity bound in the presence of a 100-fold excess of unlabelled huFN. The results are the mean of three experiments $\pm$ s.e.m.

To test the specificity of the trypomastigote binding site for huFN, inhibition experiments were done using various heterologous proteins. Incubation with thyroglobulin, ferritin, catalase, aldolase, human IgG or bovine serum albumin (BSA) had no significant effect on the binding of ^{125}I -huFN to *T. cruzi* trypomastigotes, whereas inhibition was observed on incubation with unlabelled huFN (Table 1).

The exact location of parasite-associated huFN was investigated using fluorescein-labelled antibody. Figure 2 shows the results of an indirect method using rabbit anti-human fibronectin

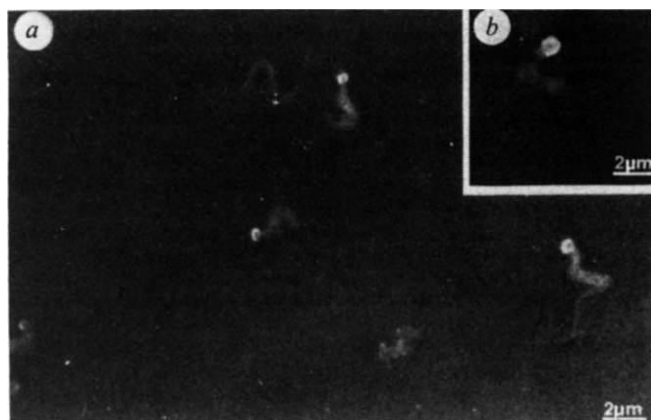


Fig. 2 *a*, Ring-shaped positive fluorescence on *T. cruzi* trypomastigote culture forms treated with huFN followed by rabbit anti-huFN antiserum. *b*, (insert), A higher magnification of the same preparation.

Methods: Rabbit antiserum to human fibronectin (anti-huFN) and its IgG fraction were obtained as reported previously²¹. The specificity of antibodies was controlled by immunoelectrophoresis²¹ and immunoblotting²² (data not shown). Trypomastigotes were washed with MEM, fixed in 0.25% glutaraldehyde in PBS pH 7.2, and dried on slides. They were then treated with the following reagents: (1) to saturate the slides and to prevent the binding of fibronectin to surfaces, they were incubated for 1 h at room temperature with BSA (Sigma) diluted 2% (w/v) in PBS followed by three washes in PBS; (2) 5 μg of human fibronectin in PBS for 1 h at room temperature, then three washes in PBS; (3) anti-huFN diluted 1:20 in PBS for 30 min at room temperature, followed by three washes in PBS; (4) fluorescein-conjugated sheep anti-rabbit γ-chain antiserum (Pasteur Institut Production, France) diluted 1:20 in PBS for 30 min at room temperature, then three washes in PBS; (5) Evan's blue diluted 1% (w/v) in PBS for 10 min. The slides were then washed twice in PBS. Control parasites were treated with normal rabbit serum instead of anti-huFN. The specimens were examined under a fluorescent microscope.

antiserum (anti-huFN) and fluorescein-labelled sheep anti-rabbit γ-chain antiserum. A ring-shaped region of fluorescence was observed on the kinetoplast region, suggesting that fibronectin binding sites are localized at the posterior end of the trypomastigote.

The amino acid composition of fibronectin from various species is very similar¹¹. Indeed, plasma and cellular fibronectin have been found by immunological criteria to be identical¹¹. The IgG fraction of the rabbit anti-huFN antiserum used in the present study reacts against rat serum and rat fibroblasts (3T3 FR) (data not shown). Thus, an *in vitro* test using 3T3 FR fibroblasts, *T. cruzi* trypomastigotes and rabbit anti-huFN IgG was done to investigate the possible involvement of fibronectin in the cell-parasite interaction. Table 2 shows the effect of rabbit anti-huFN IgG (1 mg ml⁻¹) on infection of 3T3 FR fibroblasts by Tehuantepec strain *T. cruzi* trypomastigotes; in the four independent experiments performed, marked inhibition of infection of fibroblasts was observed. The range of inhibition was 63–69%. In some experiments, either the fibroblasts or the *T. cruzi* trypomastigotes were preincubated for 1 h with either rabbit anti-huFN IgG or normal rabbit IgG at the same concentration, then washed and assayed in the culture system using Eagle's minimum essential medium (MEM) supplemented with 5% heat-inactivated fetal calf serum (FCS). In these conditions, no inhibitory effect of anti-huFN was observed (data not shown), suggesting that the constant presence of anti-huFN antibodies is necessary to inhibit infection of 3T3 FR by trypomastigotes. The inhibitory effect of anti-huFN antibodies on the infection of fibroblasts by *T. cruzi* trypomastigotes suggests that the antibodies bind to fibronectin on the fibroblast cell surface, thus preventing the trypomastigotes from attaching to the fibroblast surface. In preliminary experiments, the addition of soluble huFN to the culture medium promoted the attachment of *T. cruzi* trypomastigotes to 3T3 FR, resulting in an increase in the percentage of infected cells (data not shown).

Table 2 Rabbit anti-huFN inhibits infection of Fischer rat fibroblasts (3T3 FR) by *T. cruzi* trypomastigotes

Expt	% Inhibition
1	69
2	65
3	69
4	63

The monolayer of 3T3 FR fibroblasts was maintained as described previously with minor modifications¹⁶. 3T3 FR were cultivated at 37 °C in MEM supplemented with 5% heat-inactivated FCS in flaskette chambers. After 4 days the cultures were washed twice with MEM, then 2 ml of MEM containing 1 mg ml⁻¹ of rabbit anti-huFN IgG and 10⁶ trypomastigotes was added to one of the flaskettes; normal rabbit IgG was added to the other flaskette as a control. The flaskettes were incubated at 37 °C in a humidified atmosphere of 5% CO₂ in air. After 24 h of culture, 1 ml of MEM containing either 1 mg of rabbit anti-huFN IgG or normal rabbit IgG was added to the corresponding flaskette. Then the fibroblast cultures were incubated for 48 h at 37 °C, after which they were washed twice in MEM to remove free-swimming parasites, fixed and stained with Giemsa. In each experiment, the number of parasitized cells was quantified for 800 cells by three observers using light microscopy. In some experiments, either trypomastigotes or fibroblasts were preincubated with either anti-huFN IgG or normal rabbit IgG for 1 h, washed and then assayed in the culture system using MEM supplemented with 5% FCS. Per cent inhibition was calculated as $(1 - T/C) \times 100$ where *T* is the mean number of parasitized cells in a culture treated with rabbit anti-huFN IgG and *C* is the mean number of parasitized cells in the control culture containing normal rabbit IgG.

As in bacteria¹², intracellular protozoan parasites attach to mammalian cells in the initial phase of the infection. The mechanism involved could rely on specific recognition of cell surface components. In the case of *T. cruzi*, infection of vertebrate cells is inhibited by a range of lectins¹³ and sugars^{14,15}. It has been observed that infection of HeLa, LL-MK2 and bovine embryonic skin and muscle cells by *T. cruzi* trypomastigotes is inhibited by *N*-acetylglucosamine. It seems, therefore, that *T. cruzi* trypomastigotes bind to *N*-acetylglucosamine-containing 'receptors' on the vertebrate host surface¹⁵. The experiments reported here suggest that at least fibroblast cell surface fibronectin, the carbohydrate moiety of which is the *N*-acetylglucosamine type containing *N*-acetyl-D-glucosamine²³, could be a candidate for recognition of the cell surface by *T. cruzi* trypomastigotes. The identification of such recognition sites could lead to better understanding of the host-parasite relationships of protozoan infections and the possible control of cell invasion.

We thank J. Cornette and M. Loyens for technical assistance; and M. F. Massard for secretarial help. This investigation was supported by INSERM (U 167) and CNRS (ERA 422).

Received 16 November 1983; accepted 12 January 1984.

- Sherman, I. W. *Microbiol. Rev.* **43**, 453–495 (1979).
- Chapman, W. E. & Ward, P. A. *Science* **196**, 67–69 (1977).
- Annable, C. R. *Fed. Proc.* **31** (Abstr.), 2480 (1972).
- Miller, L. H., Dyvorak, J. A., Shiroishi, T. & Durocher, J. R. *J. exp. Med.* **138**, 1597–1601 (1973).
- Chapman, W. E. & Ward, P. A. *J. Immun.* **116**, 1284–1288 (1976).
- Jack, R. M. & Ward, P. A. *J. Immun.* **124**, 1566–1573 (1980).
- Brener, Z. A. *Rev. Microbiol.* **27**, 343–382 (1973).
- Dvorak, J. A. in *Proc. Symp. on New Approaches in American Trypanosomiasis Research* 109–120 (Pan American Health Organization, 1976).
- Hynes, R. O. & Yamada, K. M. *J. Cell Biol.* **95**, 369–377 (1982).
- Beachey, E. H. & Simpson, W. A. *Infection* **10**, 107 (1982).
- Pearlstein, E., Gold, L. L. & Garcia-Pardo, A. *Molec. cell. Biochem.* **29**, 103–128 (1980).
- Ofek, I. & Beachey, E. H. in *Bacterial Adherence* (Beachey, E. H. ed.) 1 (Chapman & Hall, 1980).
- Henriquez, D., Piras, R. & Piras, M. M. *Molec. Biochem. Parasit.* **2**, 359–366 (1981).
- Andrews, N. W. & Colli, W. *8th A. Meet. on Basic Research on Chagas' Disease*, Caxambu, Brazil, Abstr. 104 (1981).
- St Mark, J. C. & Dvorak, J. A. *Molec. Biochem. Parasit.* **5**, 333–341 (1982).
- Santiago, A. R., Afchain, D. & Capron, A. *Annls Soc. belge Med.* **61**, 369–377 (1981).
- Scatchard, G. *Ann. N.Y. Acad. Sci.* **51**, 660–663 (1949).
- Ruoslahti, E., Vuento, M. & Engvall, E. *Biochim. biophys. Acta* **534**, 210–218 (1978).
- McConahey, P. J. & Dixon, F. J. *Int. Archs Allergy appl. Immun.* **29**, 185 (1966).
- Afchain, D., Le Ray, D., Fruit, J. & Capron, A. *J. Parasit.* **65**, 507–514 (1979).
- Ouaisi, M. A., Cornette, J. & Capron, A. *Parasitology*, **88**, 85–96 (1984).
- Towbin, H., Staehelin, T. & Gordon, J. *Proc. natn. Acad. Sci. U.S.A.* **76**, 4350–4354 (1979).
- Carter, W. G., Fukuda, M., Lingwood, C. & Hakomori, S. *Ann. N.Y. Acad. Sci.* **312**, 160–177 (1978).

Recollections of physics

R.H. Dalitz

The Birth of Particle Physics. Edited by Laurie M. Brown and Lillian Hoddeson. Cambridge University Press: 1983. Pp.412. £27.50, \$44.50.

OVER the past decade, particle physics has made great progress. We now recognize quark-lepton families to be fundamental and accept that their interactions are governed by symmetry principles and the gauge mechanism. At the same time, since these gauge theories are renormalizable and calculable, we have had finite theories with acceptable properties before us for the first time and much experience has been gained concerning their mathematical nature and their physical implications.

This situation made it natural for Laurie Brown and Lillian Hoddeson to suggest that physicists should now look back over the early years during which particle physics first developed out of cosmic-ray and nuclear physics, going up to, say, the discovery of the pion in 1947. Their book consists of the contributions to the International Symposium on the History of Particle Physics which resulted, sponsored by Fermilab and held there in May 1980. It also includes two panel discussions and some additional papers, two being extensions of remarks made from the floor, the others being recollections from seven physicists not present at the conference.

Some strong impressions remain after reading through these recollections. The greatest surprise is the degree to which experiment was carried out independent of theory. Today, we are all very conscious of how in 1928 Dirac set up his relativistic equation which predicted the existence of negative energy states for electrons, and how these were interpreted as being due to positive energy particles with positive charge (particles we now know as positrons). However, the existence of the positron was not established in response to this theoretical work. The experimenters were in fact confused by the unfamiliar ideas involved in Dirac's theory and by much of the controversy it generated amongst the theoreticians, and did not whole-heartedly accept the concept of the positron. In the words of Carl Anderson, "the discovery of the positron was wholly accidental", coming about through investigations of the cosmic radiation, using new techniques, by Anderson at CalTech and Blackett and Occhialini at Manchester. This was "an age of innocence for experimental physics", as Rossi says, where fundamental exploratory experiments could still be done for little expenditure and without need for theoretical guidance. It is worth noting Serber's remark that "progress in physics depends more often on the development of experimental techniques rather than on theoretical ideas".

At the same time, the theoreticians were ignoring the signals being given by the experimenters. Anderson and Neddermeyer demonstrated quite early (by 1934) that the penetrating component of the cosmic radiation consisted equally of positively and negatively charged particles, and, a little later (1936), that the mass of these particles was not that of the proton or the electron, but lay between them. No theoretician paid attention.

After the heady experience of the birth of quantum mechanics, the theoreticians were in a revolutionary mood. As Schweber points out here, Heisenberg sought to introduce a fundamental length, Dirac used Hilbert spaces with negative norm, while Bohr said repeatedly "this is not crazy enough". Their general expectation was that quantum electrodynamics would necessarily prove inadequate in higher energy regimes. When theoretical calculation failed to agree with experiment, this was accepted almost fatalistically; in

one case at least, the disagreement was due only to numerical errors but nobody had sufficient conviction to attempt the calculation again. Oppenheimer, for example, whose influence on theoretical physics in the United States was enormous, repeatedly expressed the view that quantum electrodynamics must be wrong, despite work such as that of Weiszacker and Williams pointing out that many high-energy quantum electrodynamic processes still involved energies only of order $m_e c^2$, if the processes were viewed in the right

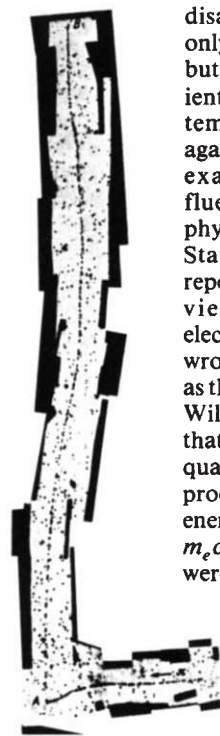
Lorentz-frame. As Serber succinctly puts it here, "Oppenheimer struggled to maintain his

lack of faith", putting forward tortured arguments for discrediting agreements found between calculation and experiment at high energies. As for new particles, there was the influence of Bohr, who followed the scholastic tradition, "do not multiply entities unnecessarily"; such an attitude suppressed speculation by Western theoreticians about new particles in the cosmic radiation, to the extent that they did not take in the message that the experimenters were so clearly spelling out.

As a result, the Japanese experience is particularly interesting. Isolation (by present-day standards) was a large factor for all in frontier physics at that time, of course. Only rarely could younger physicists manage to attend a conference. Visitors and letters were eagerly awaited, while personal contacts were of the greatest importance. But the Japanese physicists heard little more than echoes. So, as Takabayashi and Hayakawa recall, it was Yukawa, working in far-off Japan, who proposed at the end of 1934 that the nuclear forces result from the existence of a meson field, whose quanta he estimated should have mass about $150 \text{ MeV}/c^2$. Yukawa did not suffer the inhibition common to Western theoreticians of that time, who feared to propose new particles but expected major changes in the theoretical framework. As Pais remarks: "... the idea of Yukawa, that was brilliant. In the deepest sense, though, it was not revolutionary. You did not need to change any of the conceptual notions of physics".

A similar situation held for more purely theoretical questions, such as how to make sense of quantum electrodynamics, which gave infinite results in higher approximations despite good general agreement at low energies between the first approximation and experiment. Rough estimates of the finite part of the higher-order corrections, using plausible cancellations between infinities or cut-offs, were occasionally attempted, even before the Second World War (e.g. for the splitting between the $2S_{1/2}$ and $1P_{1/2}$ levels in hydrogen, where experimental discrepancy with Dirac theory was suspected already in 1937). As Lamb recalls, Oppenheimer and Hall made a suggestive estimate of this kind in 1930 with a result not far from being correct — but the estimate was disbelieved, mainly because of the fundamental doubts about quantum electrodynamics but also because these cancellations followed no consistent philosophy. As Schwinger sums it up, "The preoccupation of the majority of involved physicists was not with analyzing and carefully applying the known relativistic theory of coupled electron and electromagnetic fields but with changing it".

After Lamb and Retherford measured the splitting mentioned above (the Lamb shift) in 1947, Bethe managed at once (according to Lamb, "while he rode the New York Central railroad on the trip from



Evidence of the two-meson sequence $\pi^+ \rightarrow \mu^+$, the decay of a pion into a muon. This photomicrograph, here much reduced, appeared in the paper by C.M.G. Lattes, H. Muirhead, G.P.S. Occhialini and C.F. Powell in Nature 159, 694; 1947. R.E. Marshak and H.A. Bethe's paper which predicted the sequence was published later that year (Phys. Rev. 72, 506), as discussed by Professor Dalitz overleaf.

New York to Schenectady") to make a roughly consistent non-relativistic calculation along the above lines. Lamb and Kroll then made a relativistic calculation; as Lamb puts it, "we simply evaluated the expressions of Oppenheimer in 1930 allowing for the filled negative energy electron states". Their calculation could just as well have been made in the early 1930s, but there would have been no data to compare the result with, and no theoretician would have believed it. Even in Japan, Tomonaga had been striving to obtain finite results from quantum electrodynamics by the idea of compensation, introducing fields of unknown particles in such a way as to cancel the divergences obtained from the known interactions. However, he had realized the importance of covariant notations, and had developed methods of calculation; when news of the Lamb shift reached Japan, as a report in a weekly magazine, he was able to carry out at once the necessary relativistic calculations. In the end, it was the conservative theoreticians who proved to be correct, not the revolutionaries.

A large part of the symposium was concerned with the identification of Yukawa's meson. As Piccioni relates, a series of crucial experiments was necessary before it was realized that the penetrating particles observed in cosmic radiation did not have the strong interactions required by Yukawa. Marshak discusses the history of the two-meson hypothesis he and Bethe published in 1947 to resolve this discrepancy, namely that Yukawa's strongly-interacting meson is copiously produced in high-energy cosmic-ray interactions but decays to a lighter but weakly-interacting meson. In fact, the group of Powell and Occhialini at Bristol had already seen, and reported in *Nature* (May, 1947), the two-meson sequence $\pi^+ \rightarrow \mu^+$ in nuclear emulsion exposed to the cosmic radiation, although this was not known to Marshak at the time (June, 1947) when this idea first came to him. Communications were much slower in those days.

Brown and Hoddeson are to be congratulated on the success of their conference and on the unique contribution it has made to the history of particle physics. The recollections of the physicists involved come together in a most illuminating way, to give a vivid overview going beyond what any historian could deduce from the printed and manuscript evidence. At the time, and afterwards, probably no one had a clear picture of all that was going on. So this conference has put us further ahead, advancing our knowledge of the interrelationships of that era, providing source material for future historians and leading to a more global view of the early history of particle physics, even for those who took part in it. □

R. H. Dalitz is Royal Society Research Professor in Theoretical Physics at the University of Oxford.

Cybernetics in human affairs

Donald MacKay

Human Systems are Different.

By Geoffrey Vickers.

Harper & Row: 1983. Pp. 188. Pbk £5.95, \$12.50.

THE publication of Norbert Wiener's *Cybernetics* in 1948 created ripples of interest right across the scientific spectrum and beyond. Although he did not coin the term (it was used by Ampère in 1838 to denote "the science of government"), Wiener was the first to popularize the idea that by analysing the flow of communication and control one could discern common scientific principles that applied equally to the self-regulation of steam engines and of human communities. In particular, one could identify in feedback loops common causes of instability and misregulation — though not always, alas, cure them.

Among those who recognized the relevance of cybernetic concepts to human affairs, none was better-balanced in his appreciation than Sir Geoffrey Vickers, who died in 1982 after a long life in the practice and analysis of administrative decision making. I first met his keen mind when reviewing a symposium on "Communication" in 1955, where for sheer intellectual quality his contribution stood out from all the rest. From then on he produced a remarkable series of books with such titles as *The Art of Judgment* (Chapman & Hall, 1965) and *Freedom in a Rocking Boat* (Penguin, 1970), in which he sought to elucidate key elements in the working of human systems, and to contrast human judgement with the type of calculation that can be mechanized in a computer.

In this his last book, published posthumously, Vickers has brought together some of his most penetrating insights into the nature of human systems and the special cybernetic problems they present. Systems are nets of relations sustained through time. Open systems, interdependent with their surroundings, can be kept in being only through mechanisms of self-regulation. The constraints of interdependence, however, set limits to the possibilities of organization, making it necessary to think in ecological terms as well as those of systems engineering if we are ever to understand the conditions for their stability. When we do this, according to Vickers, the crisis of our time appears in

● Routledge & Kegan Paul have issued in paperback John C. Eccles's *The Human Mystery: The Gifford Lectures of 1977-1978*. The book first appeared in 1979, published by Springer-Verlag, and was reviewed in *Nature* 282, 160. Price of the paperback is £5.95, \$10.95.

a new and ominous light. "The threats which face the world today", he claims, are all threats to the maintenance of systemic relations on which we depend. By a threat I mean that we are no longer maintaining these relations within tolerable limits and we have at present no means in sight for doing so [p.107].

Much of the book is devoted to spelling out the nature of the threats to stability thus identified, and the implications for our society. Technology, for example, has encouraged expectations of an increasing share for all.

Yet anyone familiar with the behaviour of systems should not need telling that they breed their own limitations and that ... the management of these limitations — not their evasion but their acceptance and management — is what politics will [soon] be all about [p.147].

Again, overemphasis on the value of equality has produced a mistrust of non-reciprocal relationships (parent/child, teacher/student, master/servant)

fatal to the effective performance of a role by either party and hence to the effective working of any organization, since all depends on the trust by each in the faithful role playing of all the others [p.84].

Himself no technologist, Vickers offers no formulae or heavy theoretical models, still less any easy recipes for the rescue of our human system. He strongly urges, however, that systemic thinking should be taught and practised in schools, so that young people might grow up knowing enough about feedback loops to take it for granted that constraints must be accepted as the price of enablements.

Human systems have become very difficult for human beings to maintain. They demand from whole populations levels of understanding and tolerance seldom before found even among the few [p.177].

There are two good reasons why we scientists in particular should read and ponder this book. The first is that it powerfully challenges the hopes placed in our profession by the dominant tradition of the past two centuries.

The ideology deriving from the Enlightenment ... has aspired to realize both liberty and equality ... by a self-regulating and self-exciting process, powered by technology and directed by human "reason". ... The contemporary outcome is the reverse of what the 19th century expected. An ever more scientific world was never less predictable. An ever more technological world was never less controllable [p.xxvii].

Secondly, if Vickers is right (or even if he is not) there is an obvious need for much more thinking along these lines by minds trained in the relevant sciences and ready for the frustrations of cross-disciplinary collaboration. If a few such are stirred to action by what he has written, we may all be the better for it. □

Donald MacKay is Emeritus Professor of Communication and Neuroscience at the University of Keele.

Disease and death at Shanidar

Michael Day

The Shanidar Neandertals.

By Erik Trinkaus.

Academic: 1983. Pp.502. \$47.50, £35.

FOR some inexplicable reason the Shanidar Neanderthal remains from northern Iraq have never enjoyed the publicity that they seem to deserve. This has now been remedied by Erik Trinkaus; he has undertaken a careful study of the human bones, the results of which form the bulk of this monograph. The geology, archaeology, fauna and dating of the site are all touched upon in order to set the finds in context, but this brief treatment raises almost as many questions as it attempts to answer.

The site is a vast limestone cavern over 50m wide and 45m deep, set in a valley wall, which has given shelter to men and animals for many thousands of years. The excavations conducted in the 1950s by Solecki and Senyurek went down for 14m through five "strata" and recovered Mousterian tools, a modern fauna and nine Neanderthal skeletons, two of which were of infants. The dating of the remains quoted by Trinkaus is of the order of 60,000–70,000 years BP on the evidence of two radiocarbon dates from an upper layer and "continuous sedimentation" — tenuous grounds for any chronological opinion, as the author openly admits.

The strength of the monograph lies in the meticulous and exhaustive descriptions of the bones and teeth, and in the following attempts to assign age, sex and stature to each skeleton on the basis of anatomical observations, comparative measurements and the application of formulae. The abundance of skeletal pathology (both traumatic and degenerative) provides opportunities for speculation as to how the injuries were sustained and how the group cared for its wounded members. These aspects of the work go some way to leaven the lump of pure anatomical description and "morphometrics".

The phylogenetic considerations quoted for this group are few. The Shanidar Neandertals are said to be indistinguishable from European and Near Eastern Neandertals from other sites in terms of the face, the anterior dental dimensions and the postcranial skeleton. Not surprisingly, this also turns out to be Trinkaus's final conclusion. It may be, therefore, that we are getting some inkling of why the remains have never attracted the interest that at first sight they merit — perhaps they do not have much to contribute to our understanding of human evolution. If that is the case, then in compensation this detailed study does attempt to bring out some oblique sidelights on being a Neanderthal. It is

clear that this group of short, muscular "old" men had a tough life that played havoc with their skeletons, producing osteoarthritis from overuse (and living in a damp cave?). They also broke their ribs, feet, arms and heads from time to time but survived to show good bony healing. The most deformed skeleton shows a withered and amputated arm, possible blindness of one eye and osteomyelitis of the clavicle. Mercifully the children seem to have been quite fit when they died.

It has been suggested before that the worst-injured male was killed in a rock fall. This now seems unlikely. Trinkaus's new appraisal of the bones has prompted three possible explanations for the injuries. One is indeed a previous rock fall followed by survival, the second is cranial damage followed by hemiplegia, the third an old nerve injury that led to paralytic atrophy of the arm. Each of these explanations, and indeed the presence of severe, yet healed,

injuries on several skeletons, does provide some evidence of the behaviour of the Neanderthals from Shanidar — survival of the infirm is a behavioural trait that implies a level of social cohesion not previously identified from equivalent sites. Other sidelights on Neanderthal life in the Middle East include a suggestion of infant head-binding, intentional burial and the non-masticatory use of teeth.

Two final questions that are always asked about Neanderthals — were they taurodont and did they have a stooping bent-knee gait? The answers are yes to the first question and no to the second. The anatomical opinion given on their locomotion is that they enjoyed "normal human striding bipedal gaits" — that is, I assume, when they were not on crutches! □

Michael Day is Professor of Anatomy at St Thomas's Hospital Medical School, University of London.

Our man in Africa

Paul A. Colinvaux

The Ecological Century: A Personal Appraisal.

By E. Barton Worthington.

Oxford University Press: 1983. Pp.206. £13.50, \$27.50.

IT SO happened that the collapse of the British Empire coincided with the final emergence of ecology as a major intellectual discipline. E.B. Worthington had a career as a very British public servant through these years of change. He has left a memoir that must be of immense value to future historians; a candid statement of a career at the top while the African Empire fell and administrators learned to listen to ecologists.

Worthington began expeditioning from Cambridge in 1927. For six years he divided his time between the Cam and the African Rift or the Nile, in the company of such men as L.C. Beadle and Vivian Fuchs. His interest in fisheries, and his view of lakes as units of production, at once marked him as one of the very few at that time with both experience of Africa and scientific instincts for resource management. Julian Huxley passed his name to a proconsul of the Indian Empire (Hailey), then charged with a survey of the resources of Africa, and Worthington became secretary to that project for four years. Home again, he spent a decade as the first director of the Fresh Water Biological Association's laboratory, before returning to Africa on commission after commission. Then to the directorship of the Nature Conservancy until 1964, broken short for a last decade (1964–1974) as Secretary to the International Biological Programme, before becoming a natural choice as environmental advisor to various British industrial

consortia. All the stages of this remarkable career are set out with a frankness that illuminates the whole human system in which Worthington lived.

The opportunities given by the Empire to field naturalists were such as dreams are made of, and the young Worthington took full advantage of them. The Cambridge cachet opened frontiers, government houses, guest houses and barracks. To a generation that has known graduate students to be machine-gunned and research travellers thrown into sordid gaols for they knew not what, this is a remote past. I once caught the tail end of it myself (also on an expedition from Cambridge); on coming upon a remote African village, a person ran out to assure me that the guest house was ready and they would start cooking supper — blissful world. Worthington revelled in it and tells us of the life as if it were the most normal of existences.

The progression from Cambridge student to chairman of committees came naturally. The friends made on expeditions all rose to eminence in their respective fields, but more than all of them Worthington seems to have combined the skills of the mediator with an early grasp of the "design with nature" approach to practical ecology. This placed him in the councils of the governors and these pages will convey to future historians more than a hint of the social milieu and attitude of the British to their colonies in these transitional years.

But the defects of the old system show through clearly enough. People trained in science were extraordinarily few in number, so that the handful who were in at the start were turned to by officialdom again and again. Each appointment of Worthington's long career seems to have come about at the invitation of a friend of youth, even to the final one with IBP — across a dinner table, "how would you like to . . .".

Most glaring is the sexism of that system. When female scientists appear in the book they are "Young ladies" or "An attractive young lady". You will look in vain for reference to "Elegant gentlemen" joining the expedition and the text is definitely without "Pretty young men". During Worthington's years at the Fresh Water Biological station at Windermere, a woman started a career there whose work seems to many of us to be perhaps the most notable ever to have come from that laboratory; but her arrival is introduced with the information that "We put her on to . . .", a phrase not used for the work of any man. Worthington has honest pleasure in recounting that this woman is now a Fellow of the Royal Society. A cynic might note that this richly deserved honour came to her only after the first light airs of social change in England made some acknowledgement of women necessary. She never was granted chairs of national committees or a powerful professorship, unlike the males of her scientific cohort.

With equal strength the failure of the Imperial British to exploit their African colonies comes through. Before Worthington's generation little seems to have been done, even to survey or catalogue resources, a point made so forcibly by Correlli Barnett in *The Collapse of British Power* (Eyre Methuen, 1972). Africa for the British was a place to be governed, and the extraction of wealth was sufficient if the resulting taxes paid for the government. The men sent to govern thought of resources for hunting and fishing, even to the extent of a mischievous planting of trout in upland rivers, not of exploitation for their own or their subjects' wealth. Worthington recounts how this attitude changed only with the advent of awareness of the value of the countryside about to be handed back to its native peoples.

If the system's warts show through in Worthington's picture, it is yet clear that its picked man deserved the trust invested in him. In groping for ecological thought he was ahead of his time, already thinking in an Eltonian way about the ecosystems of African lakes in the year that C.S. Elton published *Animal Ecology*, the first modern textbook of ecology (Sidgwick, 1927). Some of the ecological assertions in the book are open to challenge, and there is some unconscious revisionism, as when he talks of his early African work in the language of energy flow, a concept then still ten years off. But he foresaw more clearly than most the need to plan for development in what now would be called an ecological way. Doubtless many of the last decisions of colonial governors were the wiser as a result of his passion. And yet it may be that this short statement of his working life, in telling of those vanished times, will be of more lasting benefit. □

Paul A. Colinvaux is a Professor in the Department of Zoology at The Ohio State University, Columbus, Ohio.

In the blood: the evolution of mankind

C.O. Carter

Blood Relations: Blood Groups and Anthropology.

By A. E. Mourant.

Oxford University Press: 1983. Pp. 146.

£15, \$20.50.

MANKIND has long been interested in the variability of his own species. The first classifications were inevitably based on physical features such as skin colour, hair colour and form, skeletal shape and facial features. These made possible the broad distinction between the Caucasoids in Europe, North Africa, south-west Asia and India, the Mongoloids in the Far East and south-east Asia, and the Negroids in Africa south of the Sahara.

But external characters were not altogether satisfactory; their inheritance is complex and some characters are shared by unrelated populations as a result of exposure to similar selective forces, for example the very dark skin found in parts of Africa, India and Melanesia. So when the ABO blood groups were discovered in 1900, and their Mendelian inheritance demonstrated in 1910 (though not fully elucidated until 1924), it was natural that analysis of their distribution should attract the attention of anthropologists. The first contribution was by the Hirsfelds at the end of the First World War. They were able to demonstrate the relatively high A frequency in Europeans and the increasing frequency of B as one moved towards eastern Europe and western and central Asia. Since then a mass of data has been collected worldwide for the ABO groups, and to a lesser extent for more recently discovered blood groups, plasma proteins, red cell enzymes and variants of the β -chain of haemoglobin.

Since the end of the Second World War, A.E. Mourant and his colleagues have systematically organized the collection of such data and published their findings and those of other workers in detail. In this short book, Dr Mourant has attempted a synthesis of all the work of preceding years.

The origins of the differences in gene frequency between populations are mutation (the original source of diversity but a rare event), natural selection and founder effect and drift. Mutation is slow to produce change; natural selection may be relatively quick; founder effect and drift operate especially when populations are small. As to the blood genes themselves, the frequencies of those for the ABO groups are relatively labile, of those for the MNS and Rhesus factors more stable. Isolated populations tend to be high in O, probably as a result of the loss of A and B

fetuses of O mothers when not counteracted by selective advantage of the A and B genes.

Mourant considers each continent in turn. Africa was the only home of man until about a million years ago, when at the *Homo erectus* stage men and women migrated to Europe, the Far East and Indonesia. Modern man, *Homo sapiens sapiens*, originating probably in western Asia, spread worldwide (including Australia but perhaps later to America) some 40,000 years ago, replacing and probably interbreeding with local less-developed varieties of man. The Sahara desert, however, became a formidable barrier to migration. It is therefore intelligible that in Africa south of the desert there are a number of "marker" genes which are rare elsewhere. These include the S^u allele and the Henshaw gene of the MNS system, the *cDe* haplotype of the Rhesus system, the J_s^a gene of the Kell system, the Fy^a allele of the Duffy system and the gene for sickle-cell haemoglobin. The selective force for the latter two is malaria; homozygotes for the Fy^a gene are almost totally resistant to *Plasmodium vivax* and heterozygotes for the sickle-cell gene have considerable resistance to *Plasmodium falciparum*.

Turning to Eurasia and Australasia, Caucasoids and Mongoloids resemble each other more than either do the Negroids. Mongoloids have less *k*, less *d* and a low but consistent frequency of the D^{β} gene of the Diego system, which is virtually absent in Caucasoids. Dr Mourant also discusses the contribution of the blood groups to several of the smaller scale questions in this area: the origin of the Basques, the Jews, the Gypsies, the Polynesians and the Ainu, and the relationship of the Australian aboriginals to tribal groups in south-west India. In addition he describes what little further is known about the relationships of blood group to susceptibility to specific diseases.

Blood Relations is an admirably clear exposition of what may be learnt from the distribution of these nuclear genes. There is much we still do not know, however; for instance Dr Mourant notes that the polymorphism of the histocompatibility genes has been studied so far only in developed countries. But here, a powerful new technique is emerging, which is outside the scope of his book. This is the study of the distribution of the polymorphisms for restriction enzyme fragments of non-nuclear (mitochondrial) DNA. This DNA is probably transmitted only by the ovum, so is inherited entirely through the maternal line and the lineage is not obscured by Mendelian segregation. □

Cedric Carter died shortly before the publication of this review. He was Head of the Medical Research Council's Clinical Genetics Unit at the Institute of Child Health, London, becoming Emeritus Professor in the University of London upon retirement. Among his works was *Human Heredity* (Pelican, 1962).